

Jukka Holopainen ja Virpi Hotti

# Jukola 2017 pilotti

## Somedataselvitys



KARELIA-AMMATTIKORKEAKOULU

# Jukola 2017 pilotti

## Somedataselvitys

KoDa – datan kokonaisvaltainen  
hallinnointi ja hyödyntäminen -hanke

*Tekijät:* Jukka Holopainen ja Virpi Hotti  
*Sivuntaitto:* Kaisa Varis  
*Kansikuva:* Venlojen viesti, Joensuu-Jukola 2017. Kaisa Varis.  
*Kustantaja:* Karelia-ammattikorkeakoulu & KoDa – datan kokonaisvaltainen  
hallinnointi ja hyödyntäminen -hanke, 2017

ISBN: 978-952-275-247-5



**Euroopan unioni**  
Euroopan aluekehitysrahasto  
Euroopan sosiaalirahasto

**Vipuvoimaa**  
**EU:lta**  
2014–2020



ITÄ-SUOMEN YLIOPISTO



# Sisällys

<b>ESIPUHE</b> .....	<b>2</b>
<b>1 JOHDANTO</b> .....	<b>4</b>
<b>2 POWER BI</b> .....	<b>7</b>
2.1 POWER BI TYÖPÖYTÄVERSION TOIMINNALLISUUDET.....	7
2.1.1 Datan lataaminen ja Query Editor .....	7
2.1.2 Näkemysten luominen .....	11
2.2 POWER BI PILVIVERSIION TOIMINNALLISUUDET.....	18
2.2.1 Pikanäkemykset.....	18
2.2.2 Q&A.....	18
2.2.3 Tilannekuva.....	21
2.3 FACEBOOK DATAN HANKKIMINEN .....	21
2.4 POWER BI NÄKEMYKSET.....	27
2.4.1 Viestien asenteiden analysointi.....	28
2.4.2 Manuaalinen luominen.....	35
2.4.3 Power BI pikanäkemykset ja Q&A .....	43
2.5 TWITTER DATA.....	49
<b>3 IBM WATSON ANALYTICS</b> .....	<b>51</b>
3.1 WATSON ANALYTICSIN TOIMINNALLISUUDET.....	51
3.1.1 Datan lataaminen ja muokkaaminen .....	51
3.1.2 Pikanäkemykset ja valmiit kysymykset .....	52
3.1.3 Näkemysten luominen .....	54
3.1.4 Tilannekuva .....	56
3.2 TWITTER DATAN HANKKIMINEN.....	56
3.3 NÄKEMYKSET TWITTER DATASTA.....	60
3.3.1 Manuaalinen luominen.....	60
3.3.2 Pikanäkemykset ja valmiit kysymykset .....	64
3.4 FACEBOOK DATAN HANKKIMINEN .....	73
3.5 NÄKEMYKSET FACEBOOK DATASTA.....	76
3.5.1 Manuaalinen luominen.....	76
3.5.2 Pikanäkemykset ja valmiit kysymykset .....	78
<b>4 GOOGLLEN ANALYTIKKATYÖKALUT</b> .....	<b>84</b>
4.1 GOOGLE ADWORDS.....	84
4.2 GOOGLE TRENDS.....	86
4.3 GOOGLE ANALYTICS.....	88
<b>5 FUTUSOMEN PALVELU</b> .....	<b>90</b>
<b>6 VIESTIEN KUVA- JA TEKSTISISÄLLÖN ANALYSOINTI</b> .....	<b>98</b>
6.1 KUVIEN SISÄLLÖN ANALYSOINTI.....	98
6.2 LAUSEIDEN SÄVYT.....	101
6.3 PERSOONALLISUUSPIIRTEET .....	104
<b>LÄHDELUETTELO</b> .....	<b>109</b>

# Esipuhe

Monipuolinen data ja sen kautta saatavat mahdollisuudet ovat viime vuosina mullistaneet liiketoiminnan kenttää. Tiedon kerääminen, analysointi ja visualisointi muodostavat nyt ja tulevaisuudessa yhden tärkeimmistä kilpailueduista yrityksille. Tiedolla johtamalla saavutetaan uusia avauksia markkinoilla ja kehitetään jo olemassa olevaa liiketoimintaa.

Mahdollisuuksista huolimatta Euroopan unionin alueella, Suomessa ja myös Pohjois-Karjalassa datan hyödyntäminen on havaittu vähäiseksi. 2016 tehdyssä Pk-yritysbarometrissä vain 3% maamme ja 2% Pohjois-Karjalan yrityksistä hyödynsi datan mahdollisuuksia liiketoiminnassaan. Luvut ovat vastaavia myös Euroopan unionin tasolla tehdyissä tutkimuksissa. Myös tulevaisuus näyttää haastavalta, sillä vain harvat yritykset suunnittelevat datan hyödyntämisen aloittamista.

Karelia-ammattikorkeakoulu ja Itä-Suomen yliopisto ovat pitkään sitoutuneet ICT-osaamisen kehittämiseen ja ammattilaisten kouluttamiseen maakunnassa. Yhteistyö on syventynyt entisestään viime vuosien aikana yhteisten hankkeiden ja koulutuksien myötä. Myös tiedon hyödyntämisen koulutusta on kehitetty vuosien varrella ja tullaan kehittämään edelleen. Tiedon ja teknologian mahdollisimman laaja hyväksikäyttö on myös osa hallitusohjelmaa sekä Pohjois-Karjalan maakunnan strategiaa. Yhdessä maakuntamme datan hyödyntämiseen erikoistuneiden yritysten kanssa oppilaitosten on näistä lähtökohdista hyvä kehittää Pohjois-Karjalan imagoa datan hyödyntämiseen liittyvän osaamisen edelläkävijä maakuntana.

Maakunnallisen kehittämisen tukemiseksi KoDa-hankkeessa oppilaitokset pyrkivät luomaan kokonaisvaltaisesta datan hallinnoinnista ja hyödyntämisestä kiinnostuneiden yritysten, oppilaitosten ja kolmannen sektorin tahojen verkoston. Verkoston tarkoituksena on tuoda avoimesti esiin alueella hyödynnettävissä olevia data aineistoja, datan hallinta ja hyödyntäminen menetelmiä sekä työkaluja. Samalla luodaan uudenlaisia monialaisia yhteistyömuotoja ja mahdollistetaan alueellisen liiketoiminnan kehittymistä.

Yhtenä potentiaalisena liiketoiminnan kehittymisen kohteena maakunnassamme ovat tapahtumat ja kolmannen sektorin toiminta. Palvelumuotoilu ja asiakaslähtöisyys ovat elintärkeitä liiketoiminnan kehittymiselle, mutta eritoten tapahtumien tyylisille elämyksellisille palveluille näiden näkökantojen huomiointi on elinehto. Tapahtumien ympärille kehittyneitä sosiaalisen median verkostoja seuraamalla voimme mahdollistaa entistä asiakaslähtöisemmät ratkaisut palveluiden suunnitteluun. Sosiaalisen median tietovarantoja voidaankin pitää asiakkaiden, tapahtuman järjestäjien ja yhteistyökumppanien näkökannasta tulevaisuuden kultakaivoksina.

Sosiaalisen median tiedon lisäksi oleellista on myös muistakin lähteistä saatava tieto tapahtumista, Saatua tietoa voidaan hyödyntää myös kansallisen ja kansainvälisen vaikuttavuuden mittaamisessa. Tiedon hyödyntäminen ja avaaminen luovat myös mahdollisuuksia uuden liiketoiminnan kehittymiselle sekä tiedon käsittely, että sen hyödyntämiselle. Samalla tuetaan yksityisen ja julkisen sektorin läpinäkyvyyttä.

Tämä asiantuntijaselvitys on toteutettu Karelia-ammattikorkeakoulun ja Itä-Suomen yliopiston yhteistyönä. Julkaisun tavoitteena on auttaa lukijaa ymmärtämään mitä

edullisia ja maksuttomia työkaluja voidaan käyttää edellä mainitun tapahtuma datan keräämiseen, analysointiin ja visualisointiin. Selvityksessä välineisiin tutustutaan Jukola 2017 tapahtuman osallistujien tietojen kautta. Näitä tietoja on kerätty verkosta kaikille avoimista lähteistä. Tavoitteena on osoittaa millaista tausta tietoa kohderyhmistä, sen mieltymyksistä ja tarpeista voidaan saada käytössä olevilla menetelmillä. Parhaimmillaan tämänkaltaista tietoa voidaan hyödyntää liiketoiminnassa esim. päätöksien, kehittämisen ja myynnin tukena.

Joensuussa,

Janne Krohns  
projektipäällikkö  
Karelia-ammattikorkeakoulu

# 1 Johdanto

Jopa heikkolaatuisesta datasta voidaan saada hyödyllisiä näkemyksiä (insights), jotka auttavat esimerkiksi määrittelemään tarkemmin mielenkiinnon kohteet. Erilaisia työkaluja hyödyntämällä saadaan esimerkiksi kuvattu, puhuttu ja kirjoitettu data muutettua käyttökelpoisiksi näkemyksiksi. Hyödynnetään enenevässä määrin luonnollisella kielellä kysymistä - on ennustettu, että vuonna 2020 jopa puolet kyselyistä tehdään luonnollisella kielellä ja jopa ääniohjatusti. Tässä selvityksessä esitellään erityisesti sosiaalisen median kanavista, *Facebook* ja *Twitter*, saatavan datan käsittelyä ja hyödyntämistä Microsoft *Power BI* ja *IBM Watson Analytics* työkaluilla – pääpaino on datasta saatavilla näkemyksillä. Kyselytoiminnallisuuksien esittelyllä halutaan rohkaista datan hyödyntämiseen. Kyselemään oppii hyödyntämällä aluksi *Watson Analytics* työkalun valmiita kysymyksiä ja kokeilemalla *Power BI*:n avainsanapohjaista Q&A toiminnallisuutta. Koska datan kerääminen ja yhdistely ovat usein aikaa vievää, selvityksessä tuodaan esille datan hankintamahdollisuus esimerkiksi Futusome Oy:n palveluna, joka koostaa eri kanavista kansalliset viestit ja tarjoaa mahdollisuuden niiden analysointiin. Lopuksi esitellään erityisesti kognitiivisten palvelujen tarjoamia mahdollisuuksia kuvien ja tekstin arviointiin.

Nykyisin hyödynnetään sosiaalisen median ja muun digitaalisen jalanjäljen jättävää toimintadataa - esimerkiksi Rikastamo-palvelu<sup>1</sup> segmentoi noin 4 miljoonaa suomalaista noin 2500 muuttujalla. Liiketoimintatiedon hallintaan tarkoitettut alustat (kuten *Power BI* tai *Tableau*<sup>2</sup>) muodostavat automaattisesti algoritmipohjaisia näkemyksiä (insights) ja erilaiset analytiikka-alustat (kuten Alteryx, Azure tai Knime<sup>3</sup>) mahdollistavat monimuotoisen datan hyödyntämisen. Myös datalähteiden omistajat, kuten *Facebook*<sup>4</sup> tai *Google*<sup>5</sup>, tarjoavat maksullisia tai maksuttomia analytiikka työkaluja käyttäjätunnuksen ja salasanan haltioille.

Tässä selvityksessä havainnollistetaan, millaisia mahdollisimman automatisoituja näkemyksiä voidaan saada ensisijaisesti kahdella kaupallisella työkalulla (*Power BI Desktop* ja sen pilviversio sekä pilvessä toimiva *IBM Watson Analytics*). *Power BI* mahdollistaa *Facebook* datasta analysoitavan datajoukon muodostamisen, joten *Power BI*:n tarjoamat analytiikkaominaisuudet esitellään ensimmäisenä (Luku 2). *IBM Watson Analytics* mahdollistaa *Twitter* datan lataamisen suoraan työkaluun, joten *Watson Analytics* työkaluominaisuudet tarkastellaan omana kokonaisuutena (Luku 3). Seuraavana esitellään Googlen tarjoamat ilmaiset analytiikkatyökalut ja niiden luomat näkemykset (Luku 4) sekä Futusome Oy:n palvelu (Luku 5). Lopuksi tuodaan esille erityisesti kognitiivisten palvelujen suomia mahdollisuuksia tekstiaineiston analysoinnille (Luku 6). Sosiaalisen median datasta haetut henkilötunnisteet on tietosuojasysteimeistä poistettu dokumentista.

---

<sup>1</sup> <http://www.aller.fi/yrittysratkaisut/rikastamo-datan-hyodyntaminen/>

<sup>2</sup> <https://www.gartner.com/doc/reprints?id=1-3RTAT4N&ct=170124&st=sb>

<sup>3</sup> <https://www.gartner.com/doc/reprints?id=1-3TISZNN&ct=170214&st=sb>

<sup>4</sup> <https://analytics.facebook.com/>

<sup>5</sup> <https://analytics.google.com>

Dokumentaatiassa kuvissa olevaan dataan ja siinä oleviin arvoihin ja kenttiin viitataan vahvistamalla. Analytiikkatyökalujen toiminnallisuuksiin, kuten valinnat, viitataan courier-fonttia käyttämällä. *Kursiivilla* viitataan esimerkiksi nimettyihin ohjeisiin tai analytiikkatyökalujen osiin, jotka eivät esiinny kuvissa. Suomen kielen yhdyssana-käytännöt jätetään tietoisesti huomioimatta, kun käytetään courier-fonttia, tehdään vahvistuksia tai kursivoitteja.

Facebook dataan kohdistettiin joukko ennalta määriteltyjä kysymyksiä ja niihin saatiin työkaluilla vastauksia (Taulukko 1).

Taulukko 1. Facebook dataan kohdennetut kysymykset ja vastaukset kuvaviittauksina työkaluittain

Kysymys	Power BI Desktop	Power BI pilvi	Watson Analytics
Mikä oli viestien yleisin sisältö?	Kuva 45.		-
Miten viestit jakautuivat määrällisesti tarkkailuajalle?	Kuva 47. Uusien viestien määrä sivulla Kuva 48. Viestien yhteismäärä ja jakauma vuosittain		Kuva 112. Uusien viestien määrä sivulla
Montako viestiä yksittäiset henkilöt liäsivät?	Kuva 49. Eri käyttäjien lähettämien viestien määrä		Kuva 123. Eri käyttäjien lähettämien viestien määrä
Mikä oli viestien asenne?	Kuva 50. Viestien asenteet		-
Minkä tyyppisiä viestejä lisättiin?		Kuva 58.	Kuva 113. Viestityyppien jakauma
Kuinka moni viesti sisälsi linkkejä tai kuvia?		Kuva 62. Viestien sisältämät linkit ja kuvat	Kuva 118. Yksilöllisten linkkien määrä Kuva 120. Linkkien määrä Kuva 121. Kuvien määrä
Mitkä olivat eniten tykätty viestit?		Kuva 63. Tykättyimmät viestit	Kuva 125. Tykättyimmät viestit

Twitter dataan kohdistettiin joukko ennalta määriteltyjä kysymyksiä ja niihin saatiin työkaluilla vastauksia (Taulukko 2).

Taulukko 2. Twitter dataan kohdennetut kysymykset ja vastaukset kuvaviittauksina työkaluittain

Kysymys	Watson Analytics
Mikä oli käyttäjien sijainti?	Kuva 82. Käyttäjien kaupungit Kuva 83. Käyttäjien sijainti valtioittain
Mikä oli käyttäjien sukupuolijakauma?	Kuva 84. Käyttäjien sukupuolijakauma
Millä kielillä twiittejä kirjoitettiin?	Kuva 85. Twiittien kielet
Miten twiitit jakautuivat määrällisesti tarkkailuajalle?	Kuva 88. Käyttäjien lähettämien twiittien määrä tarkkailuajalla
Montako twiittia yksittäiset henkilöt lähettivät?	Kuva 91. Twiittien määrä lähettäjäkohtaisesti
Minkä tyyppisiä twiittejä lähetettiin?	Kuva 93. Twiittien tyyppien jakautuminen
Mitkä olivat jaetuimmat twiitit?	Kuva 97. Jaetuimmat twiitit
Kuinka suuri osa viesteistä oli vastauksia?	Kuva 95. Vastausten määrä twiiteistä
Paljonko oli asenteeltaa positiivisia twiittejä ja paljonko negatiivisia?	Kuva 103. Negatiivisten ja positiivisten asenteiden suhde

Googlen dataan kohdistettiin joukko ennalta määriteltyjä kysymyksiä ja niihin saatiin työkaluilla vastauksia (Taulukko 3).

Taulukko 3. Googlen dataan kohdennetut kysymykset ja vastaukset kuvaviittauksina työkaluittain

Kysymys	Google AdWords	Google Trends
Kuinka paljon sanaa "jukola" haetaan Googlen hakukoneella?	Kuva 129. Avainsanojen selaaminen	Kuva 132. Hakutermin jukola tulokset
Miten Jukolaan liittyvien hakutermin kiinnostus muuttui tarkkailuajalla?	Kuva 132. Hakutermin jukola tulokset	Kuva 133. Hakutermin vertailu
Mistä muista aiheista Jukolaa hakeneet käyttäjät olivat kiinnostuneita?	Kuva 130. Hakulauseiden suosion selvittäminen	Kuva 132. Hakutermin jukola tulokset
Mistä sijainneista Jukolaa haettiin Googlen hakukoneella?		Kuva 132. Hakutermin jukola tulokset



## 2 Power BI

*Power BI* on Microsoftin analytiikkatyökalu. Työkalusta on saatavilla ilmainen kokeiluversio<sup>6</sup>, jota voi käyttää rajoitetuin toiminnoin henkilökohtaiseen käyttöön. *Pro*-versioon on saatavilla ilmainen 60 päivän kokeilujakso, jonka jälkeen se maksaa 8,40€ kuukaudessa. *Power BI Premium* puolestaan on tarkoitettu laajemman skaalan koko organisaation kattavaan käyttöön ja se hinnoitellaan tilannekohtaisesti käyttäjämäärän ja kapasiteetin käytön mukaan. Tässä raportissa kuvatut työvaiheet on suoritettu *Power BI* perusversiolla. Yksityiskohtaiset ohjeet *Power BI*:n käyttöösi liittymään ja toimintoihin löytyvät sen verkkosivuilta<sup>7</sup>. Luvussa 2.1 tarkastellaan *Power BI*:n työpöytäversion toimintoja, luvussa 2.2 pilviversioiden toimintoja, luvussa 2.3 *Facebook* datan lataamista, luvussa 2.4 ladatun *Facebook* datan näkemyksiä ja luvussa 2.5 *Twitter* datan lataamista.

### 2.1 POWER BI TYÖPÖYTÄVERSION TOIMINNALLISUUDET

*Power BI*:n työpöytäversio on suunniteltu tietojen lataamiseen, datajoukkojen muokkaamiseen sekä näkemysten manuaaliseen luomiseen. Työpöytäversiolla ei voi luoda tilannekuvia (dashboard) tulosten esittämiseksi. Luvussa 2.1.1 kuvataan datan lataaminen ja Query Editor ja luvussa 2.1.2 näkemysten luominen.

#### 2.1.1 Datan lataaminen ja Query Editor

Datajoukot *Power BI*:hin ladataan yläreunan työkalupalkin painikkeesta *Get Data* (Kuva 1). Valittavana on useita eri lähteitä<sup>8</sup> kategorioista tiedosto (*File*), tietokanta (*Database*), *Azure*, online palvelut (*Online Services*) ja muut (*Other*) (Kuva 2).

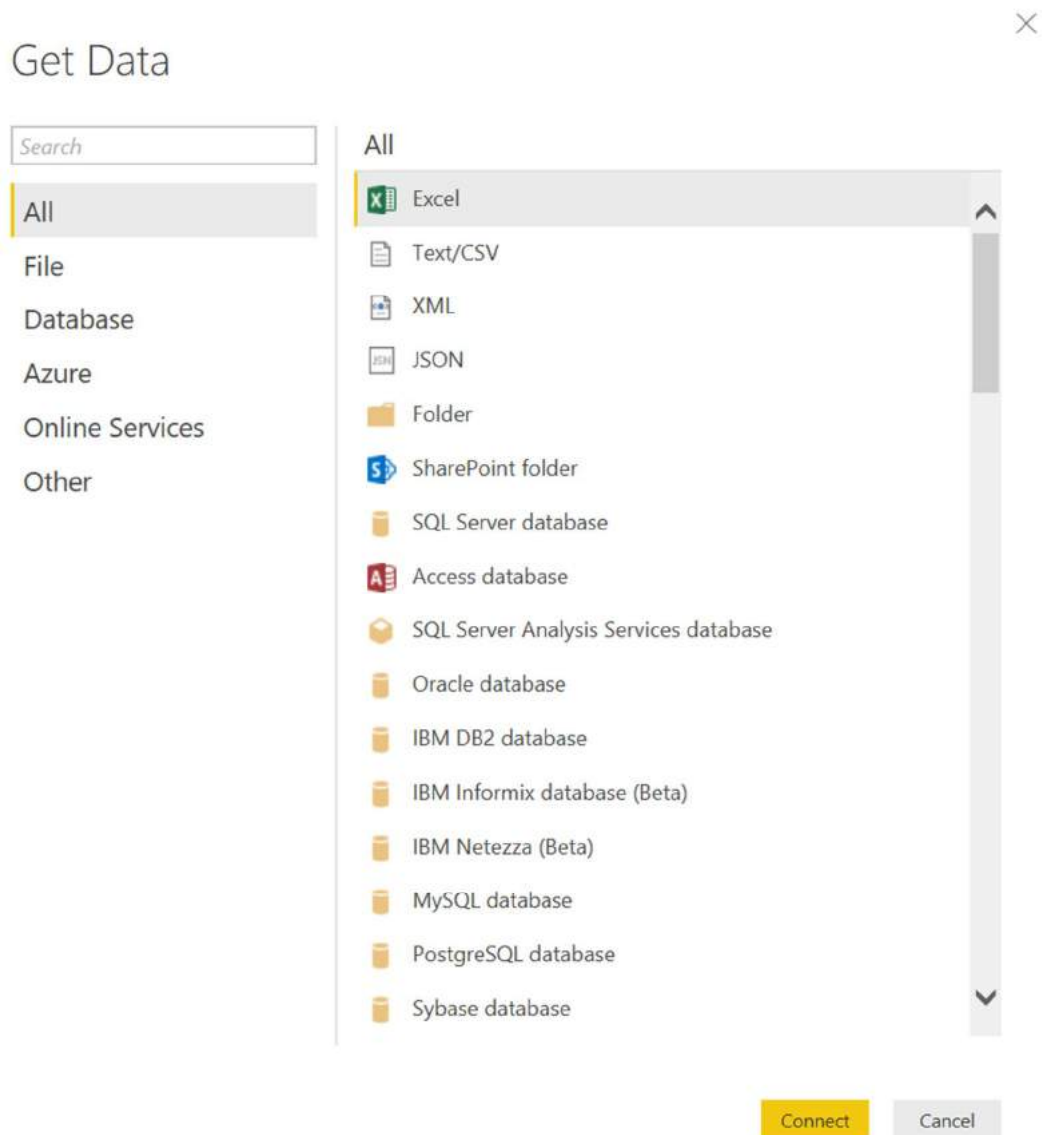


Kuva 1. Datan lataaminen

<sup>6</sup> <https://powerbi.microsoft.com/en-us/pricing/>

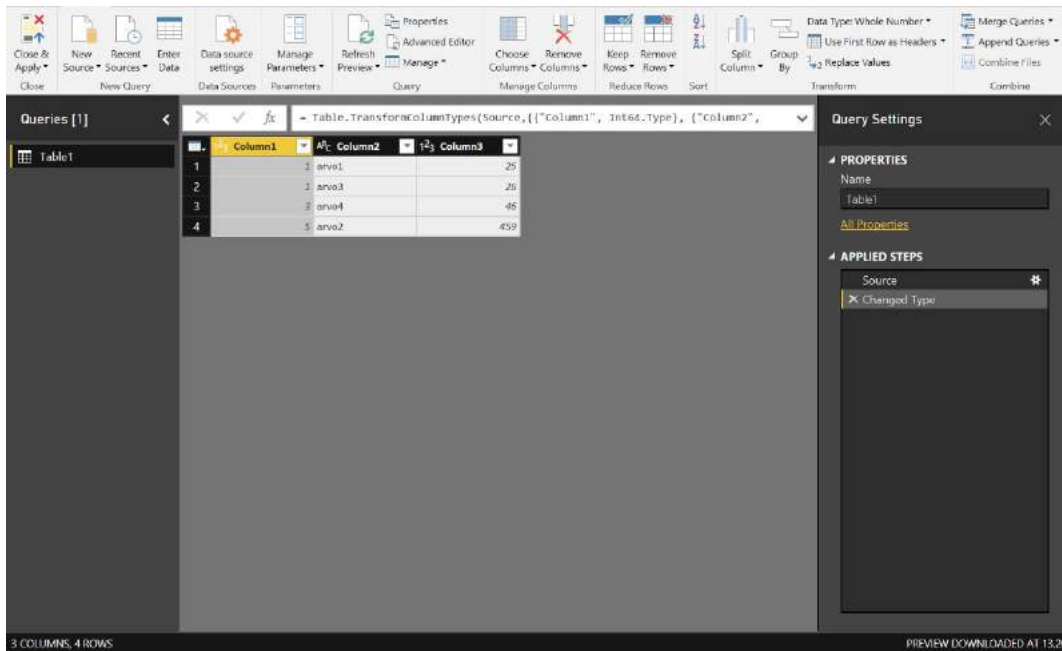
<sup>7</sup> <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-getting-started/>

<sup>8</sup> <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-data-sources/>

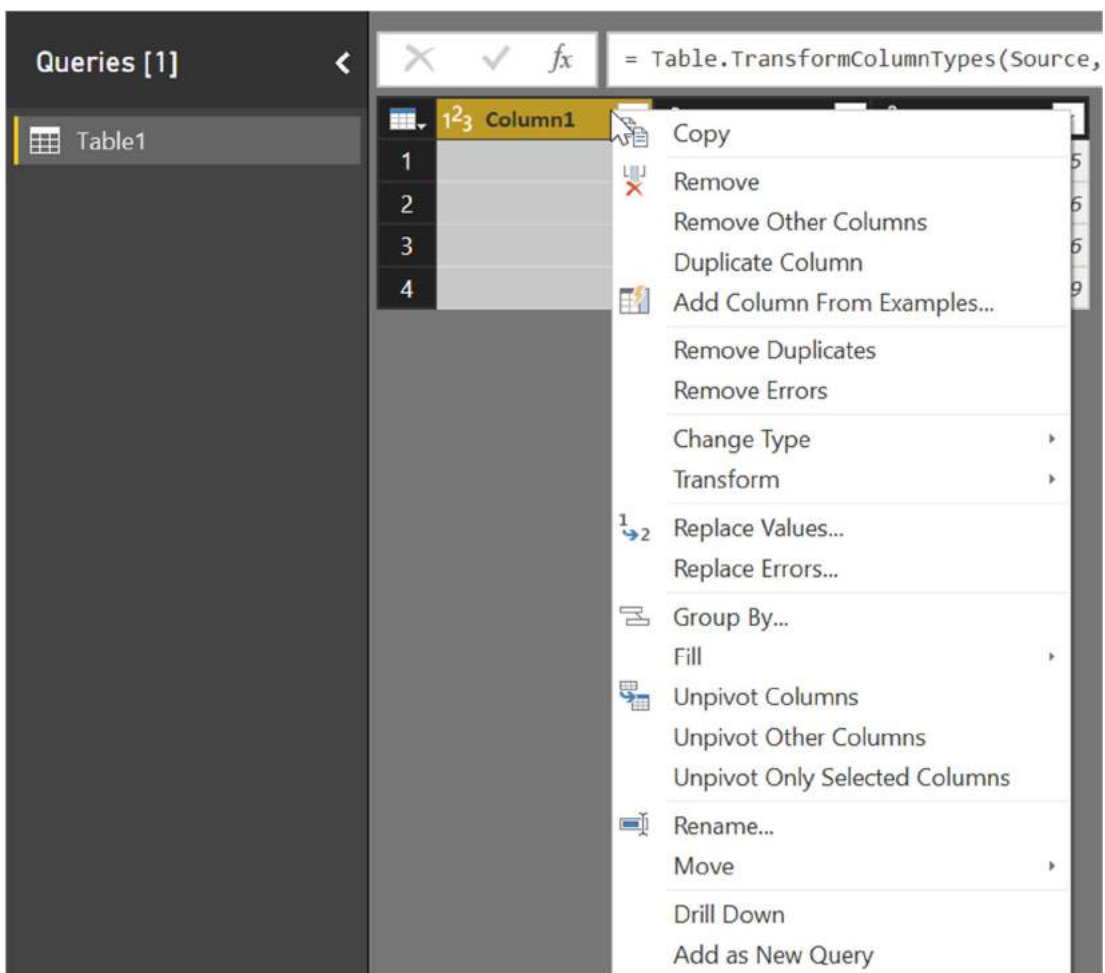


Kuva 2. Power BI Datalähteitä

*Query Editor* avautuu automaattisesti datajoukon ladattua ja on tarkoitettu kyselyjen luomiseksi sekä niiden sisällön muokkaamiseksi (Kuva 3). *Query Editorin* voi avata myös yläreunan työkalupalkin *Edit Queries* painikkeesta. Muokattavan kyselyn voi myös aukaista vasemman reunan *Query* valikosta. Yksittäisen solujen sisältöä pystyy muokkaamaan painamalla solua. Yläreunan työkalupalkki sisältää toimintoja, joilla kyselyjen dataa pystyy muokkaamaan kokonaisuutena. *Query Editorin* tarkoituksena on muokata data käsiteltävään muotoon. Tämä tarkoittaa esimerkiksi arvojen korvaamista ja sarakkeiden tyyppien muuttamista. Hiiren oikealla saraketta painamalla päästään käsiksi myös muihin työkaluihin, kuten virheiden poistamiseen ja sarakkeen kopioimiseen (Kuva 4). Mahdolliset virheet ja väärän tyyppiset sarakkeet estävät usein näkemysten luomisen ja niiden olemassaolo on hyvä tarkastaa aina. Työkalupalkin toiminnot kohdistuvat valittuna olevaan sarakkeeseen.

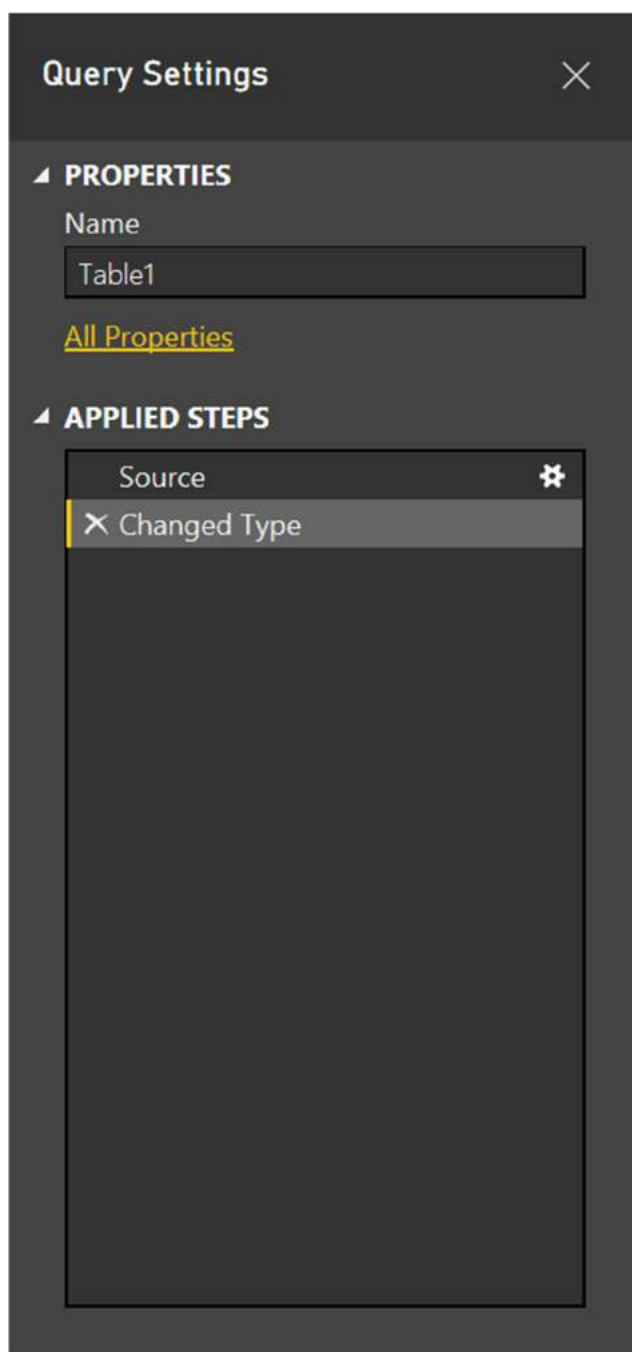


Kuva 3. Query Editor



Kuva 4. Sarakkeen muokkaaminen

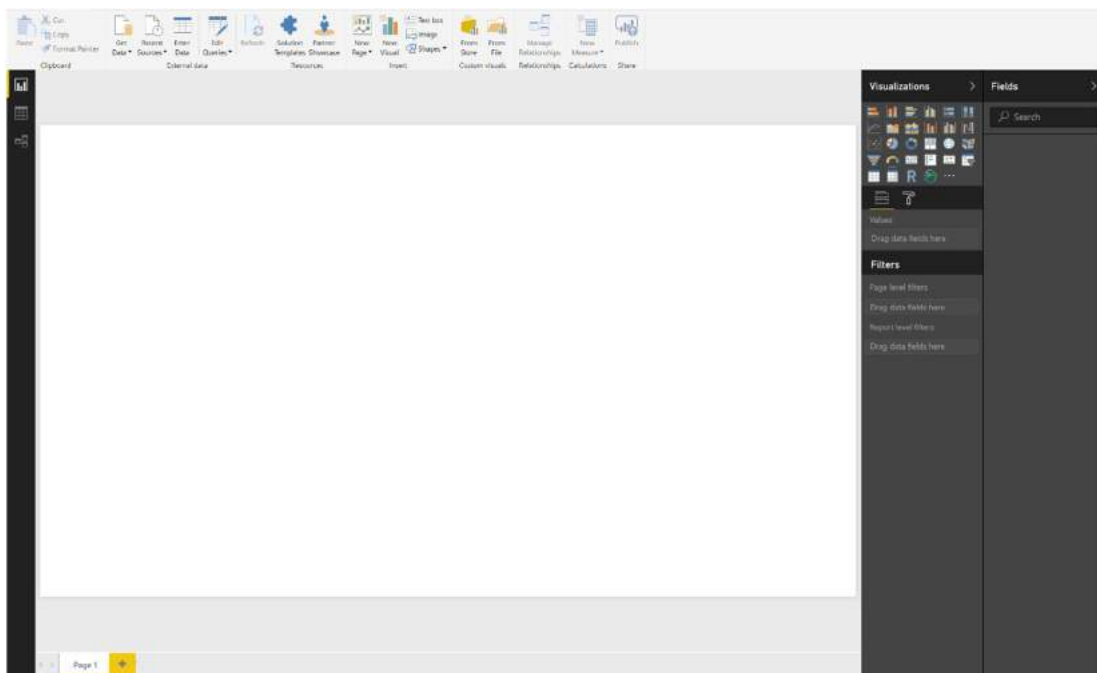
Oikean reunan `Query Settings` osiosta pystytään muokkaamaan kyselyn nimeä, sekä tarkastelemaan kyselylle suoritettuja toimintoja (Kuva 5). Jokainen suoritettu toiminto, kuten virheiden poistaminen tai tyyppimuunnokset, tulee näkyviin `APPLIED STEPS` kentässä aikajärjestyksessä. Toiminnon jälkeiseen tilaan voidaan palata painamalla sen nimeä listasta. Muutoksia voidaan myös perua poistamalla ne listasta painamalla vasemman reunan rastia.



Kuva 5. Query Settings

## 2.1.2 Näkemysten luominen

*Power BI*:n työpöytäversion aloitusnäkyman (Kuva 6) vasemmassa laidassa ovat välilehdet Report, Data ja Relationships. Näkemysten luominen tapahtuu Report välilehdessä. Jokaiselle näkemykselle voi tehdä oman välisivunsa alhaalta tai ne voi luoda samalle sivulle muiden kanssa. Data välilehdessä tietoja voi katsella *Query Editorin* tapaisessa taulukkomuodossa ja Relationships välilehdessä voi luoda yhteyksiä kyselyjen välille. Yläreunan työkalupalkista löytyvät toiminnot muun muassa datan lataamiseen ja oikean reunan työkalupalkkia käytetään näkemysten luomiseen.



Kuva 6. Power BI työpöytäversion aloitusnäkymä

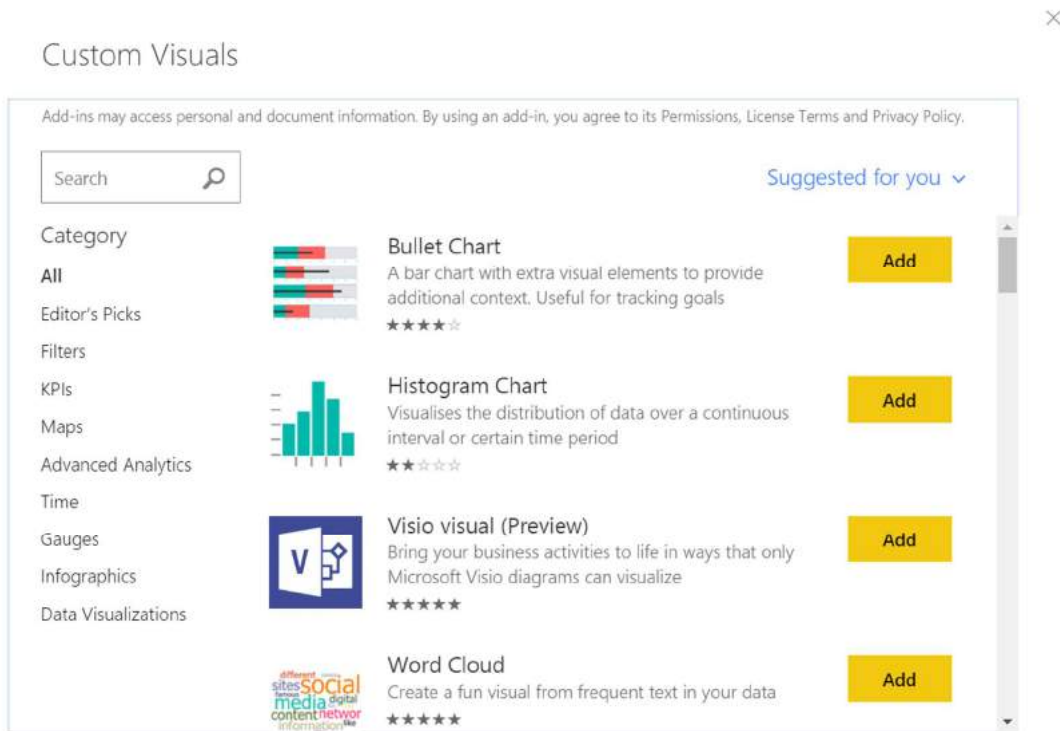
Visualisointeja voidaan lisätä oikean reunan työkalupalkista (Kuva 8) painamalla visualizations valikosta työkalua, joka lisää valitun visualisoinnin raportti-ikkunaan. *Power BI* sisältää valmiiksi 28 eri visualisointityökalua<sup>9</sup> näkemysten luomiseen. Lisää työkaluja voi ladata *Power BI*:n kaupasta painamalla työkalupalkista ... > Import from store tai valitsemalla yläreunan työkalupalkista Custom visuals osiosta From Store. Kauppa sisältää sekä ilmaisia että maksullisia visualisointityökaluja. Kaikki visuaalit eivät ole *Microsoftin* omia, vaan muiden käyttäjien tai organisaatioiden luomia, joten niiden hyödyllisyys ja helppokäyttöisyys vaihtelevat laajasti. Custom Visuaalit ovat yksinkertainen tapa luoda monimutkaisempia analyysejä ilman R kielen osaamista. Osa monimutkaisimmista visualisoinneista vaatiikin R-ohjelmointikielen asentamista toimiakseen. Huomioitavaa kuitenkin on, että R visuaalit toimivat pilvessä

---

<sup>9</sup> <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-visualization-types-for-reports-and-q-and-a/#power-bi-visualizations>

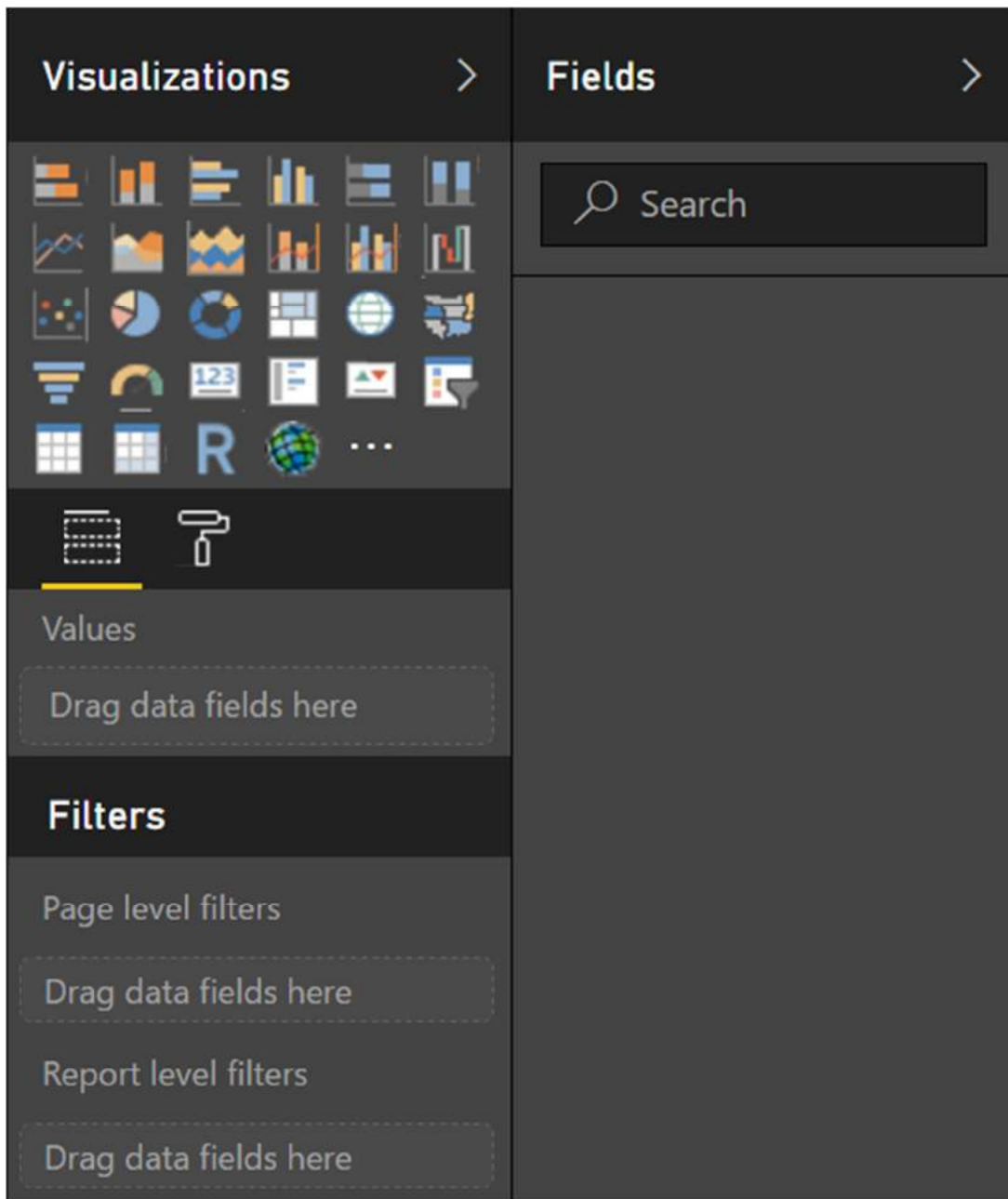
vain *Power BI Pro* versiossa, eikä niitä voi ilmaisversiolla esimerkiksi lisätä tilannekuviin.

Visualit ovat jaettu kahdeksaan kategoriaan: suodattimet (*Filters*), Key Performance Indicators (*KPIs*), kartat (*Maps*), edistynyt analytiikka (*Advanced Analytics*), aika (*Time*), mittarit (*Gauges*), infograafit (*Infographics*) ja datan visualisoinnit (*Data Visualizations*). Custom Visuaaleilla on mahdollista luoda esimerkiksi tarinoita, etsiä samanlaisuuksia klusteroimalla dataa ja etsiä datasta jos-niin yhteyksiä assosioinneilla.

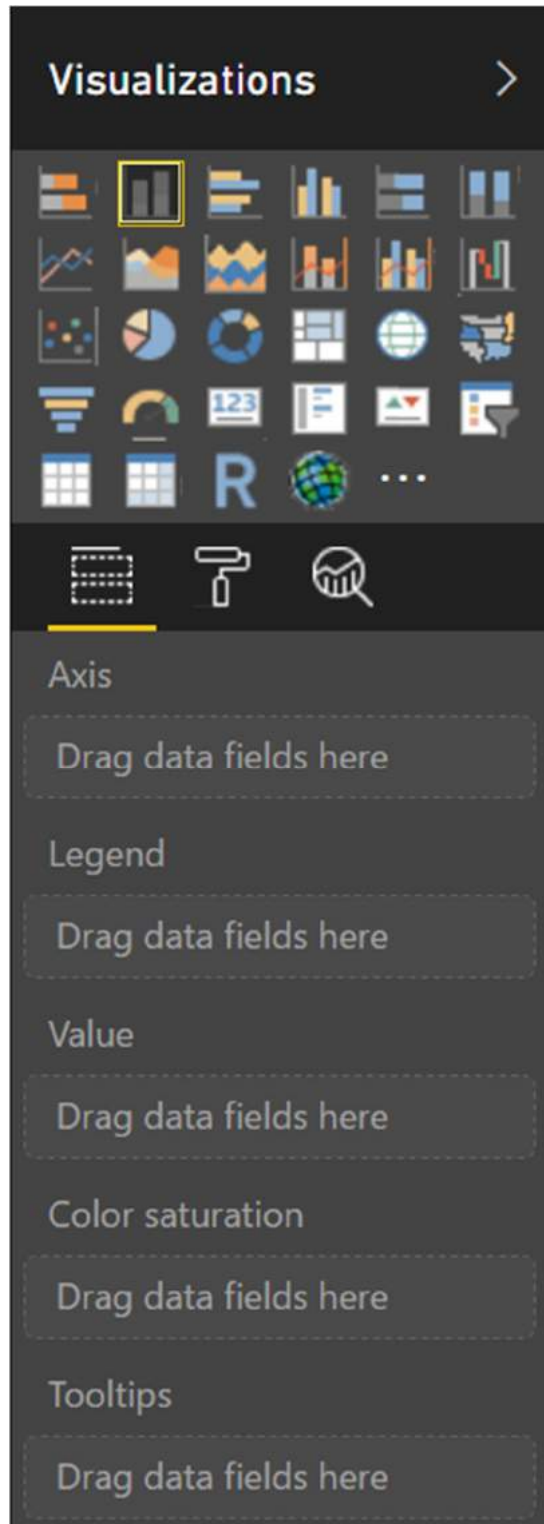


Kuva 7. Custom Visuals kauppa

Visualizations valikko on jaettu kolmeen välilehteen: *Fields*, *Format* ja *Analytics*. *Fields* välilehdessä näkyvät valitun visualisointityypin kentät, esimerkiksi diagrammin x- ja y-akselit (Kuva 9). Kenttiin lisättyjä sarakkeita pystyy järjestämään uudelleen hiiren osoittimella vetämällä. *Format* välilehdessä on mahdollista muokata visualisoinnin ulkonäköä (Kuva 10). Muun muassa värejä, otsikoita, rajoja, taustaa ja fonttikokoja pystyy vaihtamaan lähes jokaiselle visualisoinnille. *Analytics* välilehdestä voi lisätä tiettyihin diagrammeihin ennustavia osia (Kuva 11). Toiminnolla pystytään ennustamaan esimerkiksi aikasarja datan kehitystä tulevaisuudessa.

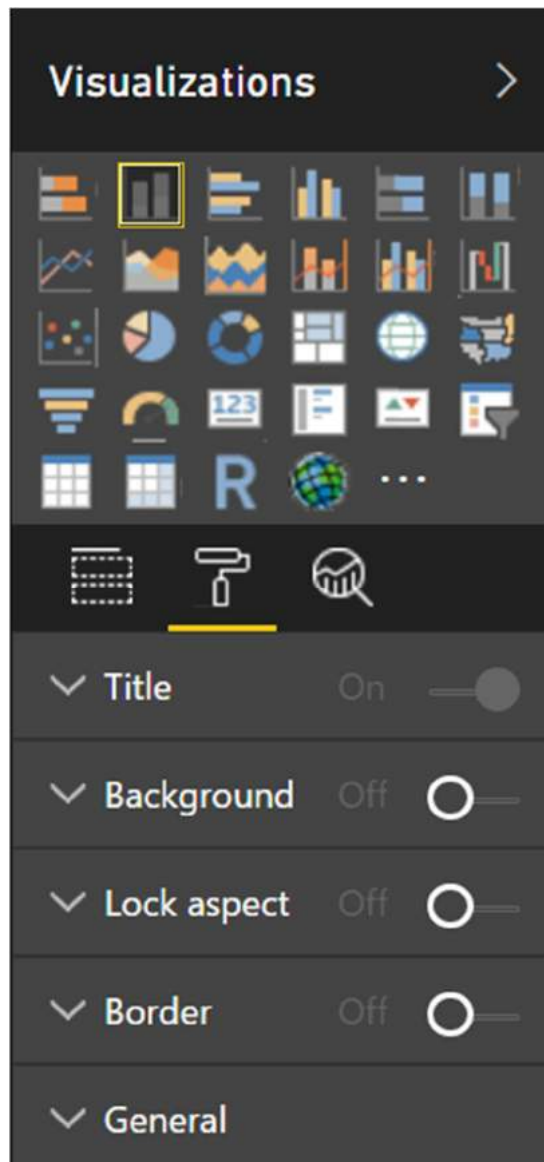


Kuva 8. Oikein reunan työkalupalkki

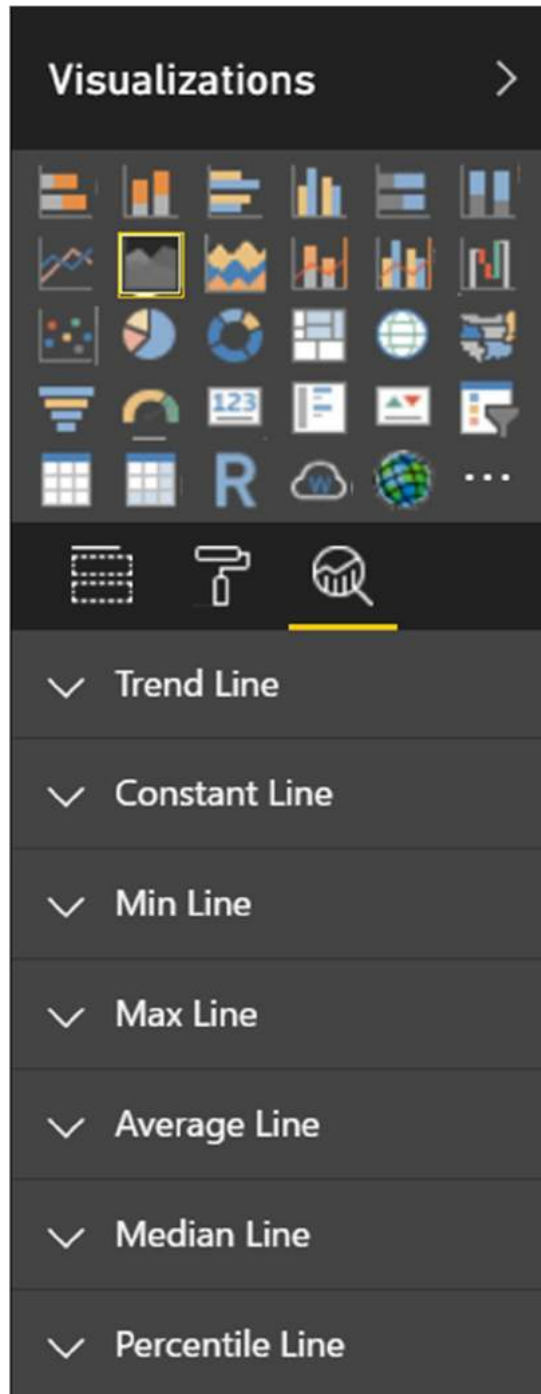


Kuva 9. Fields välilehti



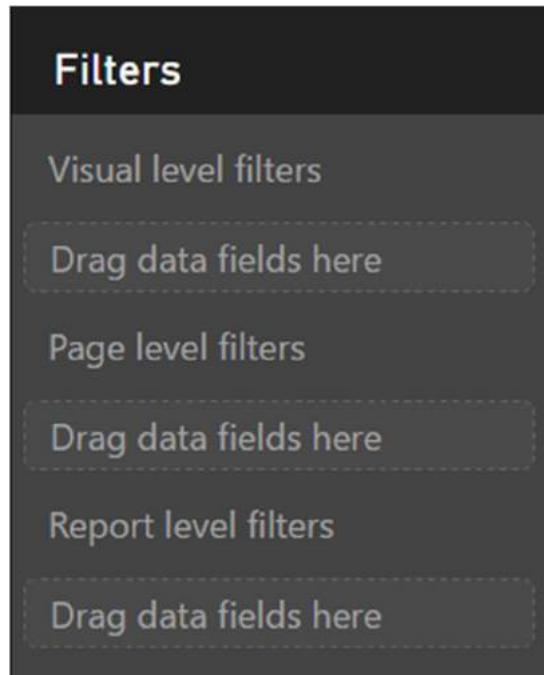


Kuva 10. Format välilehti



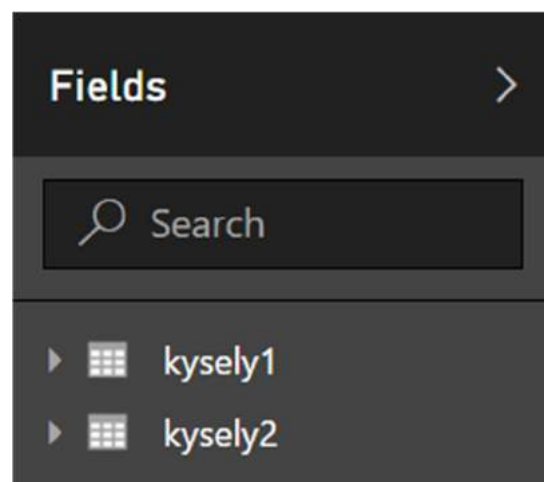
Kuva 11. Analytics välilehti

Filters osio on suodattimien luomista varten (Kuva 12). Suodattimia voi luoda kolmelle eri tasolle: Visual level filters, Page level filters ja Report level filters. Visual level filters suodatin on voimassa vain sillä hetkellä valittuna olevaan visualisointiin, Page level filters suodatin vaikuttaa raportin avonaiseen välilehteen ja Report level filters raporttiin kokonaisuudessaan. Suodatin luodaan vetämällä hiiren osoittimella kyselyn sarake Filters osioon, jonka jälkeen sen ehtoja voi säätää esimerkiksi kaaviossa näytettävänä arvoväleinä.



Kuva 12. Filters osio

Luotujen kyselyjen data tulee näkyviin `Fields` osioon. Sarakkeita voi vetää `Visualizations` valikon `Fields` välilehden kenttiin hiiren osoittimella. Kyselyjen sarakkeita voi myös etsiä hakusanoin `Search` kentästä. Sarakkeita voi myös valita painamalla niitä. Visualisoinnin voi valita ennen tai jälkeen sarakkeiden lisäämistä. Jos visualisointityyppiä ei ole etukäteen valittu `Visualizations` valikosta, luo *Power BI* sen automaattisesti vastaamaan parhaiten valittujen sarakkeiden arvoja.



Kuva 13. Fields osio

## 2.2 POWER BI PILVIVERSION TOIMINNALLISUUDET

*Power BI*:n pikanäkemykset ja Q&A toiminnallisuus ovat käytettävissä vain sen pilviversiossa<sup>10</sup>. Vaikka raportteja voi muokata myös pilviversiossa, on se huomattavasti helpompaa työpöytäversiolla sen nopeuden ansiosta. Lisäksi *Query Editori* on käytettävissä vain työpöytäversiolla.

Pilviversio sisältää neljä osiota: DASHBOARD, REPORTS, WORKBOOKS ja DATASETS. DATASETS osiossa voidaan luoda uusia raportteja työpöytäversion tapaan. REPORTS osiossa voidaan muokata aiemmin työpöytäversiolla luotuja raportteja sekä luoda niihin uusia näkemyksiä. WORKBOOKS osiosta löytyvät Microsoft Excelistä *Power BI*:hin julkaistut työkirjat. DATASETS osiossa pystytään luomaan uusia näkemyksiä ja raportteja pilviversioon ladatuista datajoukoista samalla tavalla kuin työpöytäversiolla. Luvussa 2.2.1 kuvataan pikanäkemykset, luvussa 2.2.2 Q&A toiminnallisuus ja luvussa 2.2.3 tilannekuvat.

### 2.2.1 Pikanäkemykset

Pikanäkemyksiä (quick insights) voi luoda pilviversiossa datajoukot (DATASETS) osiosta. Datajoukkoa on myös mahdollisuus optimoida pikanäkemyksiä varten esimerkiksi piilottamalla sarakkeita<sup>11</sup>. Pikanäkemyksien etsiminen aloitetaan painamalla DATASETS osiossa datajoukkoa hiiren oikealla ja valitsemalla *Quick insights*. Kun työkalu on luonut näkemykset, ne saa näkyviin valitsemalla samasta kohdasta *view insights*, joka aukaisee pikanäkemykset (Kuva 56). Pikanäkemykset auttavat löytämään nopeasti datasta uusia näkemyksiä sekä luomaan nopean kokonaiskuvan datajoukon suhteista. Hyödyllisiä pikanäkemyksiä voi kiinnittää tilannekuvaan ruudun yläkulman nastasta (*Pin*), jolloin ne tulevat näkyviin uutena ruutuna jo luotujen rinnalle.

Pikanäkemyksiä voidaan etsiä myös jo valituista pikanäkemyksistä. Painamalla pikanäkemyksiä se avautuu tarkasteluun ja edelleen painamalla oikean yläkulman painiketta *Get Insights* sivun oikeaan laitaan avautuu näkemykseen liittyviä pikanäkemyksiä.

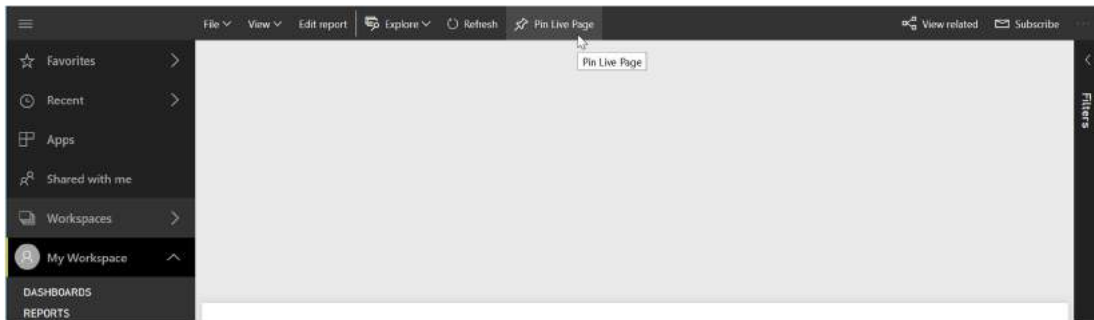
### 2.2.2 Q&A

Jotta raportista voi kysyä kysymyksiä Q&A toiminnolla, tulee se kiinnittää tilannekuvaan yläpalkin painikkeesta *Pin Live Page* (Kuva 14) tai kiinnittämällä datajoukon pikanäkemyksen tilannekuvaan. Kysymyksiä voi kysyä kaikista datajoukoista, joiden näkemyksiä on lisätty samaan tilannekuvaan.

---

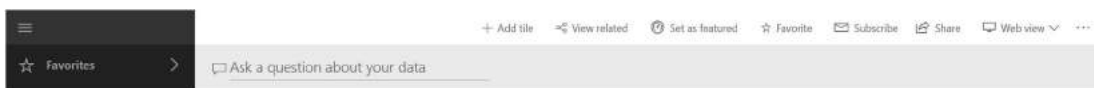
<sup>10</sup> <https://powerbi.microsoft.com/en-us/landing/signin/>

<sup>11</sup> <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-auto-insights-optimize/>



Kuva 14. Tilannekuvan luominen

Tilannekuva näkyessä (Kuva 15) on yläreunassa tekstikenttä `Ask questions about your data`, jonka avulla datajoukosta voi kysyä kysymyksiä luonnollisella kielellä englanniksi. Tekstikenttää painamalla työkalu ehdottaa valmiiksi datajoukon sarakkeiden nimiä, jotka esiintyvät datajoukossa. Kentän voi valita klikkaamalla ehdotusta tai kirjoittamalla sen käsin.



Kuva 15. Kysymysten esittäminen

Q&A hyväksyy englannin kielisiä avainsanoja<sup>12</sup> ja tarjoaa valittavaksi erilaisia kysymyksen sopivia vaihtoehtoja. Avainsanojen luokat ovat kokoaminen (aggregate), artikkelit (Articles), tyhjä ja boolean arvo (Blank and Boolean), vertailut (Comparisons), konjugaatiot (Conjugations) päivämäärät (Dates), suhteelliset päivämäärät (Relative dates), alueen yhdenmukaisuus (Equality (Range)), arvojen yhdenmukaisuus (Equality (Value)), nimet (Names), Prononymit (Pronouns), kyselykomennot (Query commands), alue (Range), ajat (Times), järjestys (Top N), visualisoinnin tyypit (Visual types) ja suhteet (Wh (relationship, qualified)). Taulukossa 4 on kuvattu avainsanaluokat, niiden yleisimmät avainsanat ja esimerkkikysymys kuvitteellisesta datajoukosta, joka sisältää sarakkeet vuosi, sukupuoli, nimi, kaupunki, sijointi ja aika. Kysymyksiä esittäessä ei tarvitse käyttää kysymyssanoja. Kysymyksen muodostamaa visualisointia voi muokata ennen kiinnittämistä tilannekuvaan samoilla toiminnoilla kuin näkemysten manuaalisessa luomisessa.

<sup>12</sup> <https://powerbi.microsoft.com/en-us/documentation/powerbi-service-q-and-a-tips/>

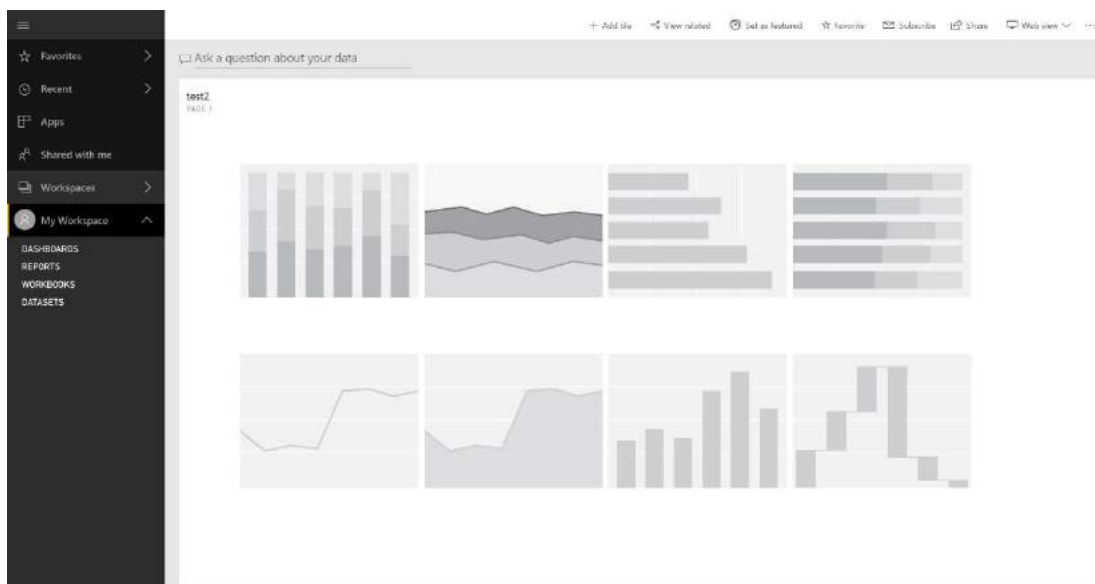
Taulukko 4. Kysymystyyppien avainsanat ja esimerkkikysymykset

Luokka	Avainsanat ja esimerkkikysymys
Aggregates	total, sum, amount, number, quantity, count, average, most, least, fewest, largest, smallest, highest, biggest, maximum, max, greatest, lowest, littlest, minimum, min "Amount of rows"
Articles	a, an, the "The amount of rows"
Blank and Boolean	blank, empty, null, prefixed with "non" or "non-", empty string, empty text, true, t, false, f "Amount of blank rows"
Comparisons	vs, versus, compared to, compared with "Amount of rows in year 2016 compared to year 2017"
Conjugations	and, or, each of, with, versus, &, and, but, nor, along with, in addition to "Nimi and sukupuoli"
Contractions	didn't, haven't, he'd, he's, isn't, it's, she'll, they'd, weren't, where'll, who's, won't, wouldn't "Nimi of who's kaupunki was not Helsinki"
Dates	day, week, month, year, quarter, decade, MonthName, Days 1-31 "Nimi from last month"
Relative dates	today, right now, current time, yesterday, tomorrow, the current, next, the coming, last, previous, ago, before now, sooner than, after, later than, from, at, on, from now, after now, in the future, past, last, previous, within, in, over, N days ago, N days from now, next, once, twice "Nimi from later than 2015"
Equality (Range)	in, equal to, =, after, is more than, in, between, before "Nimi where sijoitus is between 10 and 20"
Equality (Value)	is, equal, equal to, in, of, for, within, is in, is on "Nimi where sijoitus is 100"
Names	datajoukkojen sarakkeiden ja solujennimet "Which Nimi are named Pekka"
Pronouns	he, him, himself, his, she, herself, her, hers, it, itself, its, they, their, them, themselves, theirs, this, these, that, those "Nimi and their sukupuoli"
Query commands	sorted, sort by, direction, group, group by, by, show, list, display, give me, name, just, only, arrange, rank, compare, to, with, against, alphabetically, ascending, descending, order "Nimi sorted by sijoitus"
Range	greater, more, larger, above, over, >, less, smaller, fewer, below, under, <, at least, no less than, >=, at most, no more than, <=, in, between, in the range of, from, later, earlier, sooner, after, on, at, later than, after, since, starting with, starting from, ending with "Nimi where sijoitus is greater than 10"
Times	am, pm, o'clock, noon, midnight, hour, minute, second, hh:mm:ss "Nimi where aika is less than one hour"
Top N	top, bottom, highest, lowest, first, last, next, earliest, newest, oldest, latest, most recent, next "Latest nimi"
Visual types	kaikki Visualizations valikon visualisointityypit "Amount of kaupunki as a map"
Wh (relationships, qualified)	when, where, which, who, whom, how many, how much, how many times, how often, how frequently, amount, number, quantity, how long, what "How long was the average aika"

### 2.2.3 Tilannekuva

Tilannekuvassa näkyvät kaikki sinne lisätyt näkemykset omina ruutuinaan (tile). Raporttityökalulla luodut raporttisivut näkyvät kukin yhtenä tiilenä ja voivat sisältää siis useampia näkemyksiä. Ruutuja voi järjestellä ja niiden kokoa muuttaa vetämällä hiiren osoittimella. Muita kuin näkemysruutuja, esimerkiksi kuvia, tekstiä tai videoita, voi lisätä yläreunan `Add tile` painikkeesta. Pro-versiossa valmiin tilannekuvan voi jakaa `Share` painikkeella.

Yksittäisen raportin sisältämät näkemykset ovat automaattisesti interaktiivisia keskenään. Esimerkiksi valitsemalla hiiren oikealla taulukosta vuosi 2016, muuttuvat myös raportin muut näkemykset esittämään valitun vuoden arvoja. Sanapilvi näyttää kyseisen vuoden yleisimmät viestien sanat, viestien määrä näyttää vuoden viestit ja viivadiagrammi viestien jakauman päivittäin. Vaikutuksia voi myös muokata raporttityökalun `Visual interactions` toiminnolla. Toiminnoilla on kuitenkin vaikutus vain saman raportin visualisointeihin, eli raporttityökalulla luotujen visualisointien manipuloinnilla ei ole vaikutusta esimerkiksi `Q&A` toiminnallisuudella luotuun kuvaajaan, joka kiinnitetään omana ruutunaan.

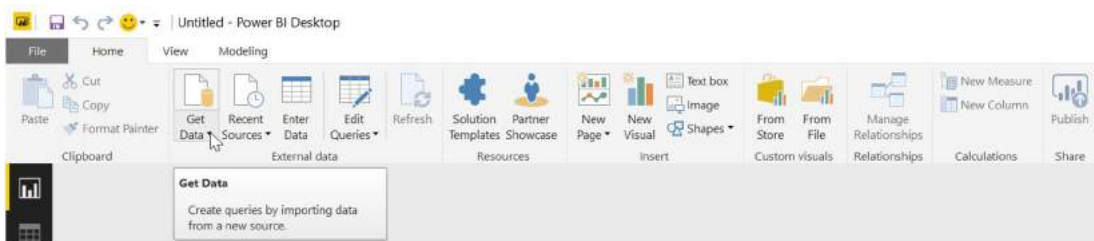


Kuva 16. Tilannekuva näkymä

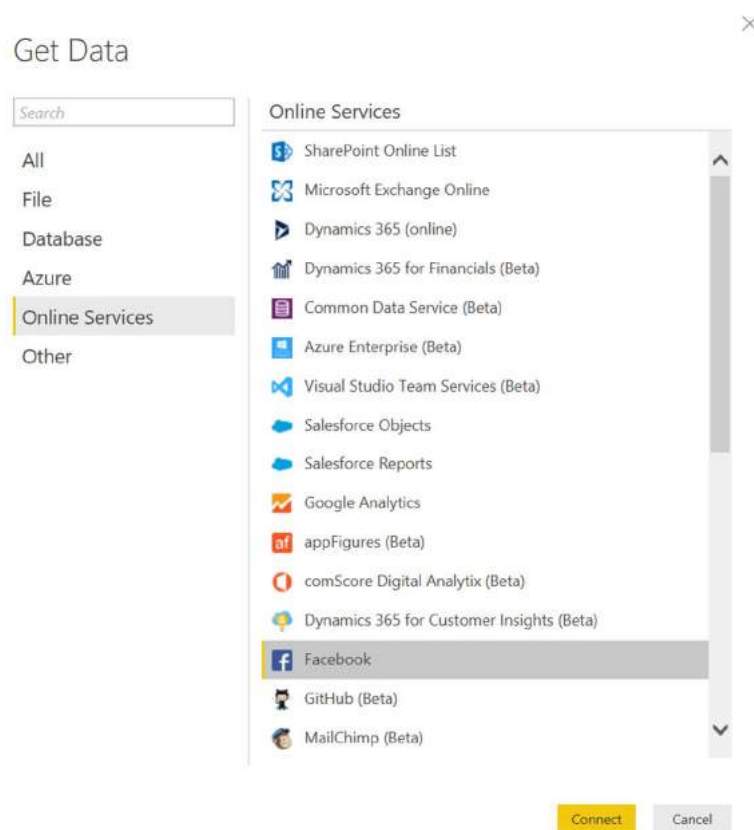
## 2.3 FACEBOOK DATAN HANKKIMINEN

*Power BI* tukee suoraan *Facebookia* tiedonlähteenä. Painetaan työkalupalkista `Get Data` (Kuva 17), joka aukaisee uuden ikkunan tiedonlähteen valinnalle (Kuva 18). *Facebookin* valinnan jälkeen määritellään haettavaksi haluttu sisältö. Ensimmäiseen tekstikenttään kirjoitetaan haun kohde ja toiseen kenttään valitaan sisältö. Jukolan viestin virallinen *Facebook*-sivu on [facebook.com/jukolanviesti](https://www.facebook.com/jukolanviesti), joten kohteena on ryhmän tunnus jukolanviesti.

Yhteys (Connection), eli haettava sisältö, voidaan valita ennalta määrätyistä luokista. Facebook päivittää API:aan säännöllisesti, joten automaattinen kommenttien hakeminen ei välttämättä aina onnistu, jos Power BI ei tue muutoksia. Valitaan yhteydeksi None, joka ei luo automaattista yhteyttä vaan antaa mahdollisuuden määrittää yhteyden jälkeenpäin Query Editorissa (Kuva 19). Ensimmäisen kerran Facebookiin yhdistettäessä työkalu kysyy Facebook kirjautumistunnuksia, joita tarvitaan yhteyden luomiseen (Kuva 20). Tunnukseksi käy esimerkiksi oma henkilökohtainen Facebook käyttäjä. Sisäänkirjautumisen jälkeen Query Editor aukeaa automaattisesti (Kuva 21).



Kuva 17. Get Data painike



Kuva 18. Tiedonlähteen valitseminen



## Facebook

"Me", user name or Object ID

jukolanviesti

Connection

--None--

[Learn more](#)

OK

Cancel

Kuva 19. Facebook sisällön valitseminen

The screenshot shows a web browser window with the URL `https://www.facebook.com/v2.8/dialog/oauth`. The page header features the Facebook logo and the text "Facebook". The main content area contains the following text: "Kirjaudu sisään ja ala käyttää Facebook-käyttäjättiliäsi sovelluksella **Microsoft Power BI**." Below this, there are two input fields: "Sähköpostiosoite" (email) with the value "facebooktest@airmail.cc" and "tai puhelinnumero:" (or phone number). A "Salasana:" (password) field is shown with masked characters. There is a checkbox labeled "Pidä minut sisäänkirjautuneena sovellukseen **Microsoft Power BI**". Below the checkbox is a blue button labeled "Kirjaudu sisään" and a link "Unohditko käyttäjättilin?". At the bottom, there is a green button labeled "Luo uusi käyttäjättili".

Kuva 20. Sisäänkirjautuminen

The screenshot shows the Query Editor interface. The top bar displays the query: `Facebook.Graph("https://graph.facebook.com/v2.8/jukolanviesti")`. Below the query, a table shows the results of the query:

name	value
name	Jukolan viesti
connections	success
id	111804412223323

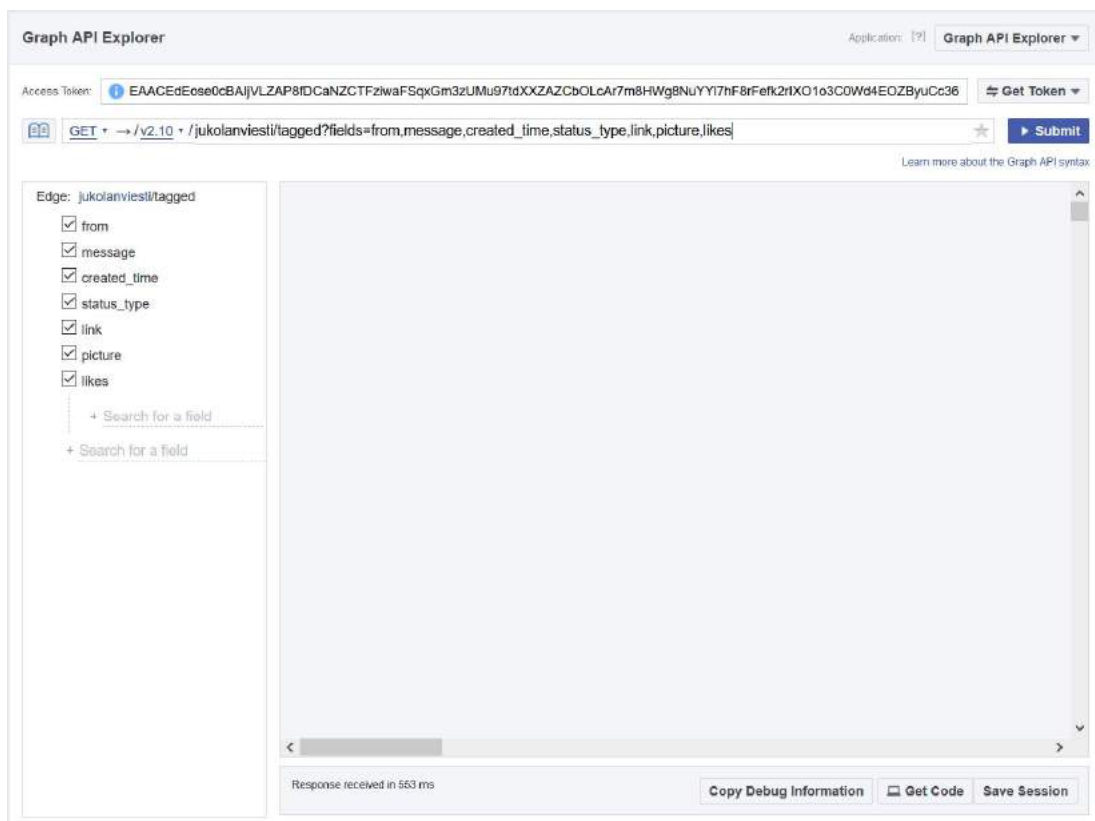
Kuva 21. Query Editor

*Query Editorissa* pystytään luomaan yhteys haluttuun osioon *Facebook* sivulla. *Power BI* käyttää automaattisesti API:n versiota 2.8, mutta sitä voi vaihtaa helposti kaavaa muuttamalla. Kaavan tekstikentässä on valmiina Jukolan viestin *Facebook* sivun yhteys:

```
= Facebook.Graph("https://graph.facebook.com/v2.8/jukolanviesti")
```

Kaavaa voidaan muokata *Facebookin Graph API ohjeen*<sup>13</sup> avulla. Kaavoja voi myös testata *Facebookin verkkosivujen Graph API Explorer* (Kuva 22) työkalulla<sup>14</sup> ennen niiden ajamista *Power BI:ssä*. Tavoitteena on analysoida sivun vierailijoiden viestejä. API:n mukaisesti vierailijoiden kirjoittamat viestit saadaan `tagged` yhteyden avulla. Viesteistä otetaan `fields` komennon avulla tiedot esimerkiksi viestin lähettäjistä, viestin tekstistä, lähetyksajasta, viestin tyypistä, mahdollisista linkeistä ja kuvista sekä tykkäysten määrästä. Viesteistä voidaan myös hakea niiden lähettäjän sijainti, mutta tämä on epäluotettava tieto, koska se sisältää sijainnin vain, jos käyttäjä on sen itse viestiinsä lisännyt. Suurin osa viesteistä ei sisällä paikkatietoja. Haettavan sisällön aikavälillä voi hallita `since` ja `until` parametreillä. *Facebook* sallii ajan rajaamisen vain enintään 35 päivän jaksolta, joka on liian lyhyt tähän analysointitarkoitukseen. Vaihetaan API:n versioksi uusin (eli 2.10) ja lisätään tarvittavat kentät API-ohjeen mukaisesti sekä rajataan haettujen sivujen määrä API:n sallimaan sataan `limit` parametrillä, jolloin kaavaksi muodostuu:

```
= Facebook.Graph("https://graph.facebook.com/v2.10/jukolan-  
viesti/tagged?fields=from,message,created_time,status_type,link,pic-  
ture,likes&limit=100")
```



Kuva 22. Facebook Graph API Explorer

<sup>13</sup> <https://developers.facebook.com/docs/graph-api>

<sup>14</sup> <https://developers.facebook.com/tools/explorer/>

Power BI haki Facebookista 788 riviä hakukriteereitä vastaavia tietoja eli viestejä (Kuva 23). Haetut sarakkeet ovat viestin lähettäjä (from), luontiaika (created\_time), viestin tyyppi (status\_type), viestin id (id), objektin linkki (object\_link), tykkäysten määrä (likes), viestin sisältämä linkki (link) ja viestin sisältämä kuva (picture).

Osa haetuista tiedoista sisältää useamman kentän, jonka vuoksi ne näkyvät *Query Editorissa* tauluina (Table). Taulut tulee purkaa, että niiden sisältöä voidaan analysoida. Tykkäyksen määrä saadaan painamalla likes sarakkeen yläkulman nuolta ja valitsemalla *Aggregate* valikosta *Count of id* (Kuva 24). Sarakkeen uudeksi nimeksi vaihtuu *Count of likes.id*. Viestin lähettäjän kertova from sarake saadaan näkyviin jakamalla sen sisältö kahteen eri sarakkeeseen valitsemalla sen yläkulman nuolesta *Expand* ja valitsemalla *name* sekä *id* (Kuva 25). Sarake jakautuu sarakkeiksi *from.name* ja *from.id*. Lopuksi tulee tarkastaa sarakkeiden datan tyyppi ja tarvittaessa korjata se oikeaksi työkalupalkin *Data Type* valikosta (Kuva 26). Oikea sarakkeiden datatyyppi on välttämätön näkemysten muodostamisessa. Sarakkeen *created\_time* tyyppiä vaihdetaan *Date/Time*, joka muuttaa sen ensin tarkaksi ajaksi ja sen jälkeen uudestaan vain *Date*, joka poistaa tarkan kellonajan. Sarakkeen *count of likes.id* tyyppiä vaihdetaan *Whole number*.

	from	message	created_time	status_type	id	object_link
1	Record	Jaukko: live -file pre-warming!!! Jukolan viesti: esä muuttanut Isovervas...	2017-07-08T11:21:32+0000	mobile_status_update	1745698019019013_1911298685792878	Record
2	Record	Sari Anttonen on yksi Suomen parhaista raissaunnistajista, joka etene...	2017-06-30T06:44:07+0000	shared_story	1616246098639686_1871066663157627	Record
3	Record	Dur 100 Moods From Finland team want to legendary Jukolan viesti to...	2017-06-29T03:46:45+0000	shared_story	94612669066_10154738130374007	Record
4	Record	Ollinme ensimmäistä kertaa tänä vuonna Jukolan viestin yhteistyökam...	2017-06-27T15:46:00+0000	shared_story	195754037175639_1381571195261578	Record
5	Record	Ecco il percorso che il nostro Paolo ha compiuto al lancio della Jukola 20...	2017-06-26T20:19:04+0000	added_photos	1925646894318386_1932934490256292	Record
6	Record	Jukolan Viesti				
6	Record	Jukolan viesti -tunnelmia kuukoosteena	2017-06-25T19:01:32+0000	added_video	186027334866677_1023016541167740	Record
7	Record	Hyvää juhannusta kaikille!	2017-06-23T18:30:40+0000	shared_story	186027334866677_1021636231305779	Record
		Jukolan viestin tunnelmia voit käydä lukemassa blogissa				
		<a href="http://veski.karjatanrasti.blogspot.fi/2017/06/jukola-jouheu-jukolassa.html">http://veski.karjatanrasti.blogspot.fi/2017/06/jukola-jouheu-jukolassa.h</a>				

Kuva 23. Kaavalla haetut tiedot

Search Columns to Aggregate

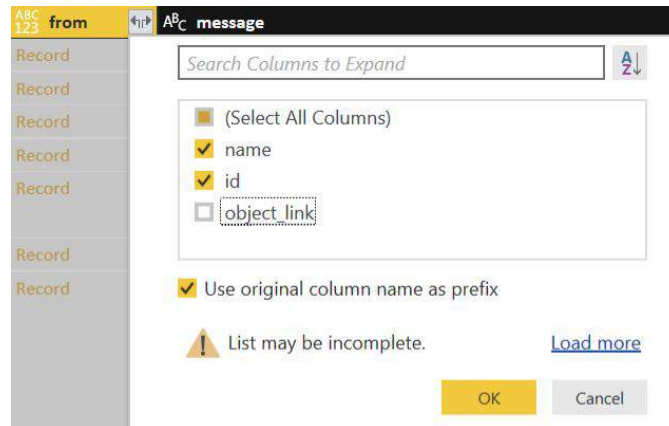
Expand  Aggregate

- (Select All Columns)
- # Count of id
- # Count of object\_link
- # Count of name

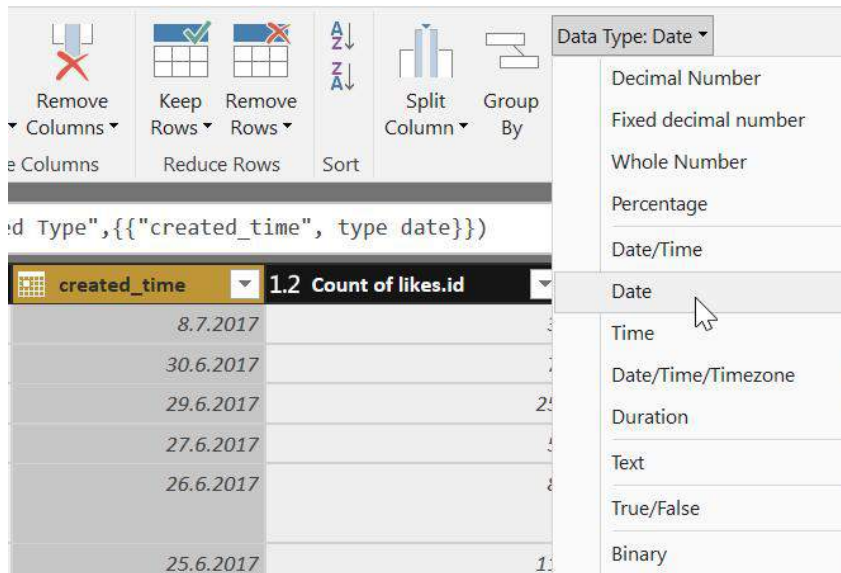
Use original column name as prefix

OK Cancel

Kuva 24. Tykkäyksen määrä

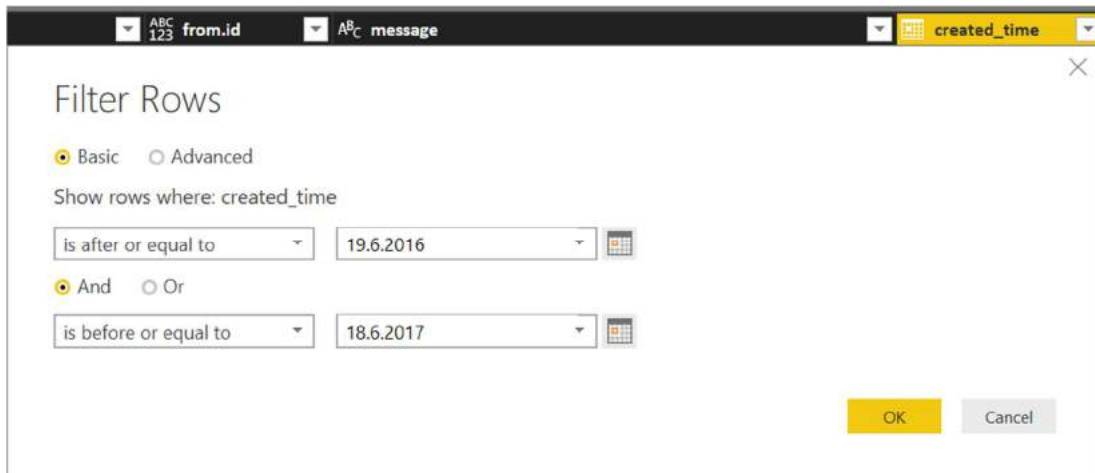


Kuva 25. Sijainnin nimi



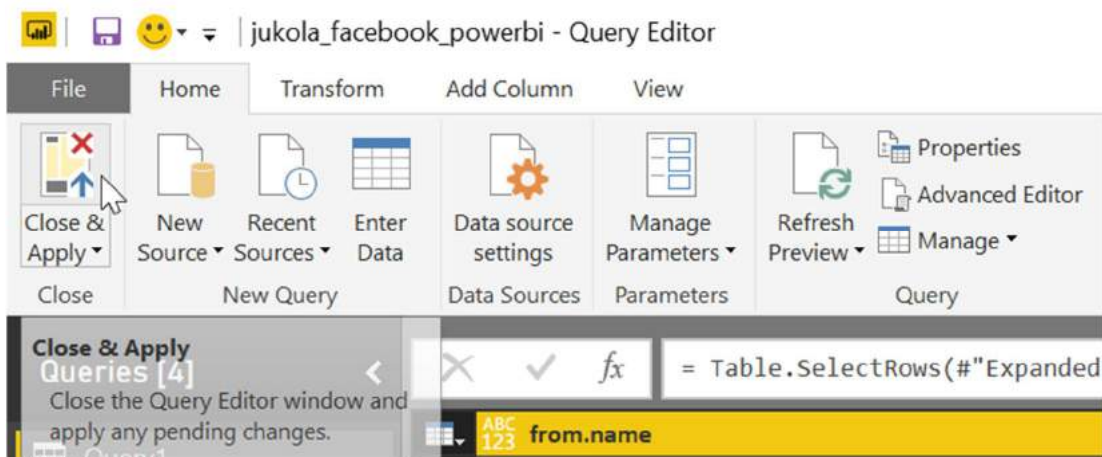
Kuva 26. Datatyyppin vaihtaminen

Kysely sisältää viestit koko Facebook sivun olemassa olon ajalta. Poistetaan ylimääräiset viestit painamalla created\_time sarakkeen nuolta ja luomalla suodatin painamalla Date filters > Custom filter. Luodaan suodatin valitsemaan viestit, jonka päivämäärät ovat tarkastelujakson sisällä (Kuva 27). Suodattamisen jälkeen jäljelle jää 188 viestiä.



Kuva 27. Suodattimen luominen

Data on nyt haettu ja tarvittavat muokkaukset tehty. Seuraavaksi data voidaan ladata käsiteltäväksi näkemysten muodostamista varten. *Query Editor* voidaan sulkea ja muutokset tallentaa painamalla yläreunan työkalupalkista *Close & Apply* (Kuva 28). *Power BI* lataa nyt kaavan mukaiset tiedot *Facebook API*:n kautta.



Kuva 28. Kyselyn muutosten tallentaminen

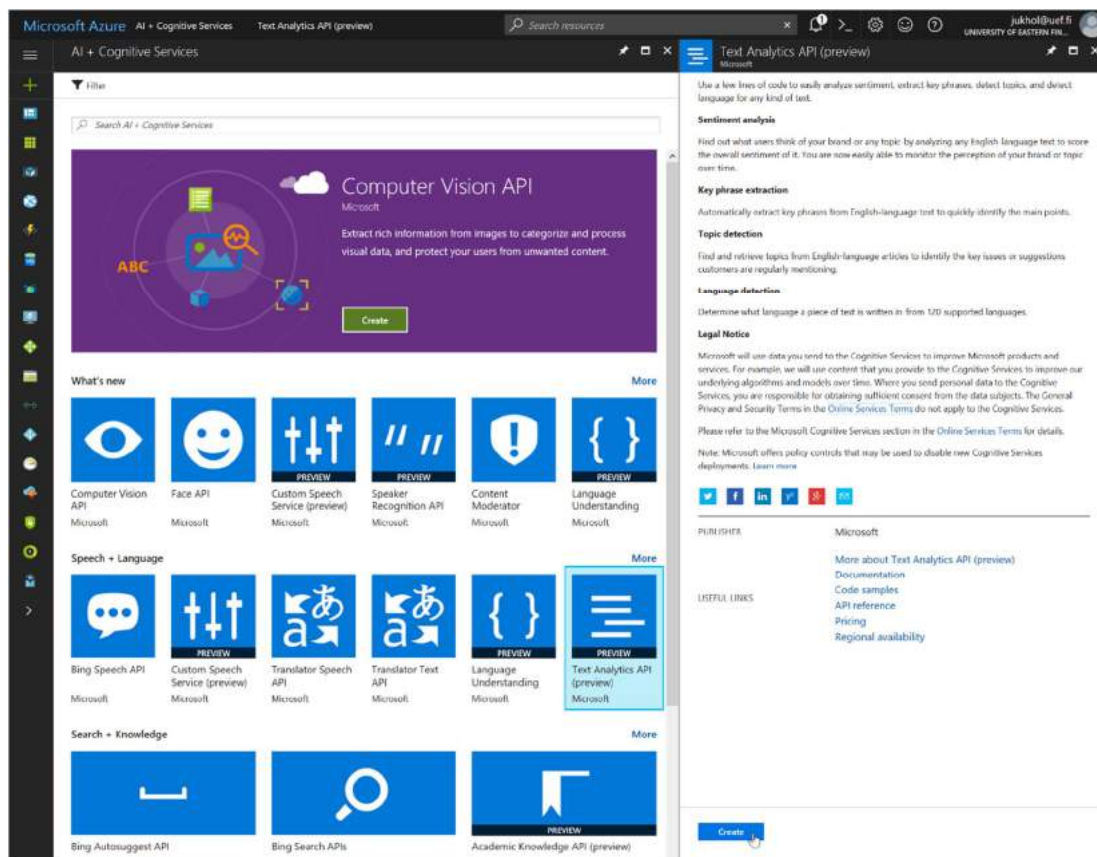
## 2.4 POWER BI NÄKEMYKSET

Tässä luvussa kuvataan näkemysten luominen *Power BI*:lla *Facebook* datalle. Luvussa 2.4.1 kuvataan asenteiden (sentiments) analysointi *Azure*n avulla, luvussa 2.4.2 näkemysten manuaalinen luominen työpöytäversiolla ja luvussa 2.4.3 näkemysten luominen pikanäkemyksen ja Q&A:n avulla sekä tilannekuvan luominen.

## 2.4.1 Viestien asenteiden analysointi

Viestien sisältöä voidaan analysoida tarkemmin Microsoft *Azuren*<sup>15</sup> kognitiivisten palveluiden<sup>16</sup> avulla. *Azuren Text Analytics API Preview* kognitiivinen palvelu kykenee havaitsemaan tekstistä sen asenteet, avainsanat (keywords) ja aiheet (topics). Rekisteröityminen Azureen on ilmaista ja myös palvelun saa kokeiluun ilmaiseksi. *Text Analytics API Preview* palvelu tukee täysin vain Englantia, mutta myös suomenkielisistä viesteistä saadaan niiden asenteet<sup>17</sup>.

Kognitiiviset palvelut ja edelleen *Text Analytics API* löytyy *Azuresta* navigoimalla sivupalkista *More services > Cognitive Services > Text Analytics API (preview)*. Palvelu luodaan painamalla *Create* (Kuva 29). Luotu palvelu tulee näkyviin tilannekuva näkymään (Kuva 30).

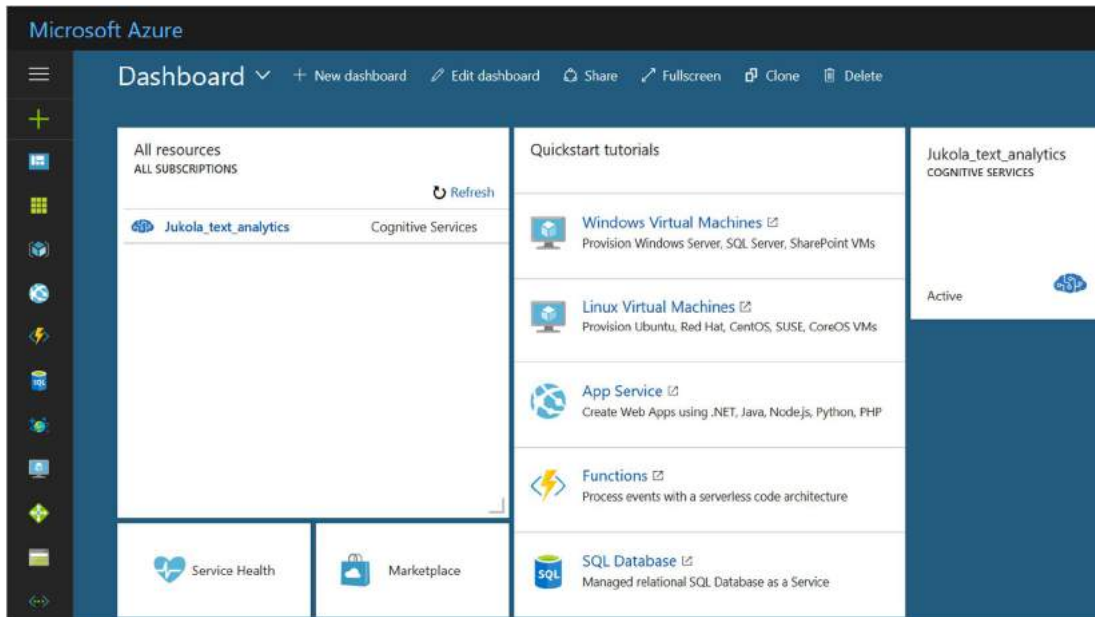


Kuva 29. Palvelun luominen

<sup>15</sup> <https://azure.microsoft.com/en-us/free/>

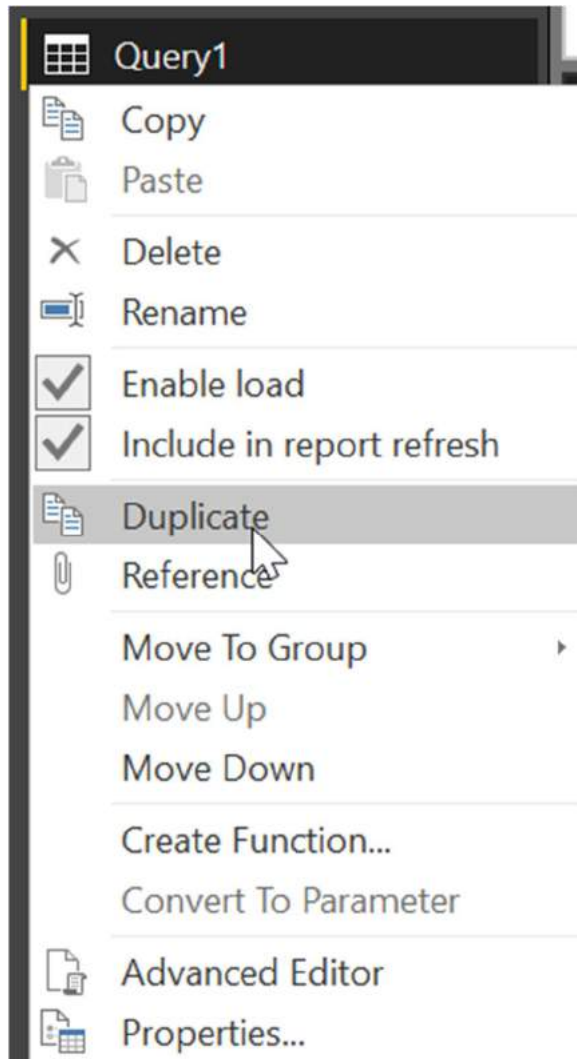
<sup>16</sup> <https://azure.microsoft.com/en-us/services/cognitive-services/>

<sup>17</sup> <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/overview#supported-languages>



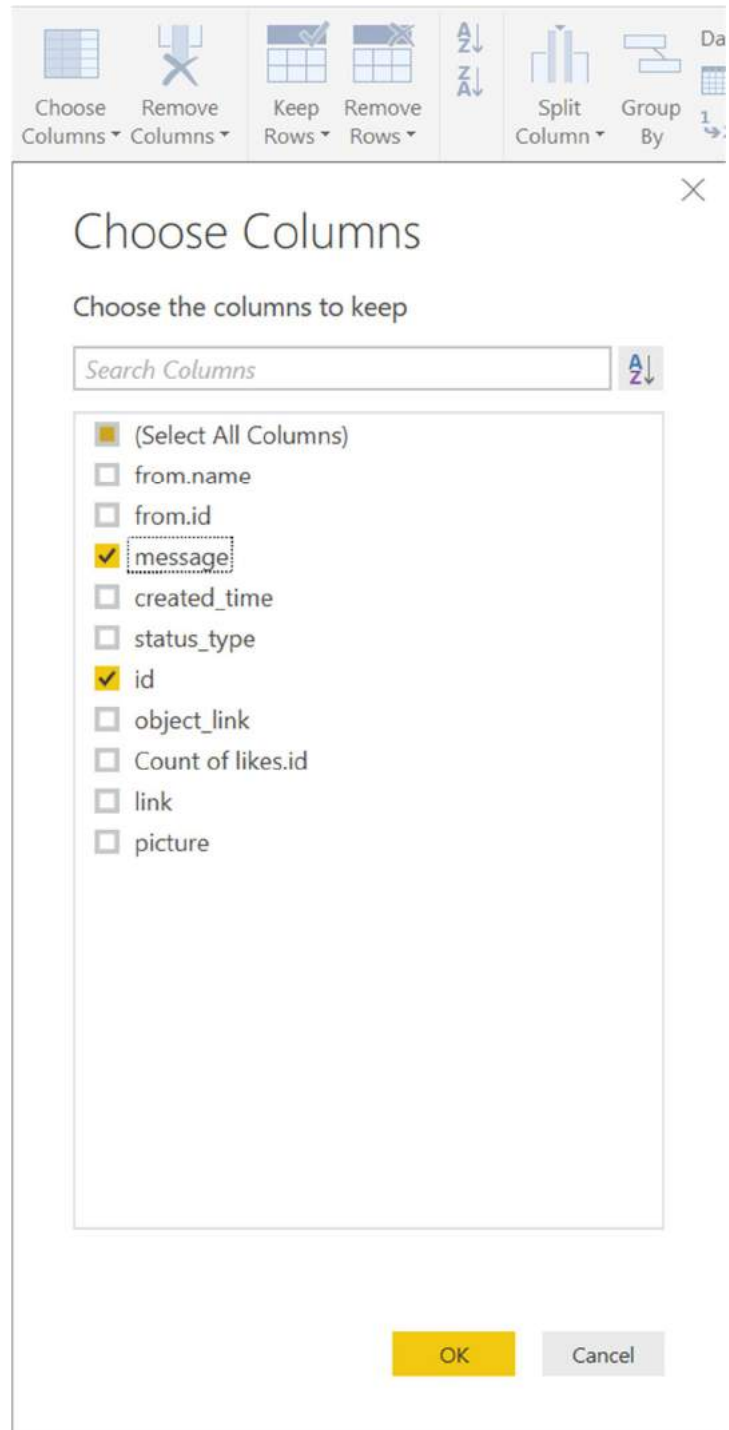
Kuva 30. Luotu palvelu

Asenteiden hakemiseksi täytyy palata *Query Editoriin*. Luodaan ensimmäisestä kyselystä toinen identtinen valitsemalla siitä *Duplicate* (Kuva 31). Uudesta kyselystä valitaan säilytettäväksi sarakkeiksi vain *message* ja *id* valitsemalla yläreunan työkalupalkista *Choose Columns*. Nimetään uudelleen sarake *messages* sarakkeeksi *text*, joka on vaatimuksena *Text Analytics API:n* käyttämiseen



Kuva 31. Kyselyn duplikaatti





Kuva 32. Sarakkeiden valitseminen

API:n käyttäminen onnistuu M kielellä. Luodaan tyhjä kysely valitsemalla yläreunan työkalupalkista *New Source > Blank Query*. Avataan kyselyn editori valitsemalla yläreunan työkalupalkista *Advanced Editor* (Kuva 34). Scripti *Text Analytics API:n* käyttöä varten:

```
(Source) =>
let
    JsonRecords = Text.FromBinary(Json.FromValue(Source)),
    JsonRequest = "{ \"documents\": " & JsonRecords & " }",
```

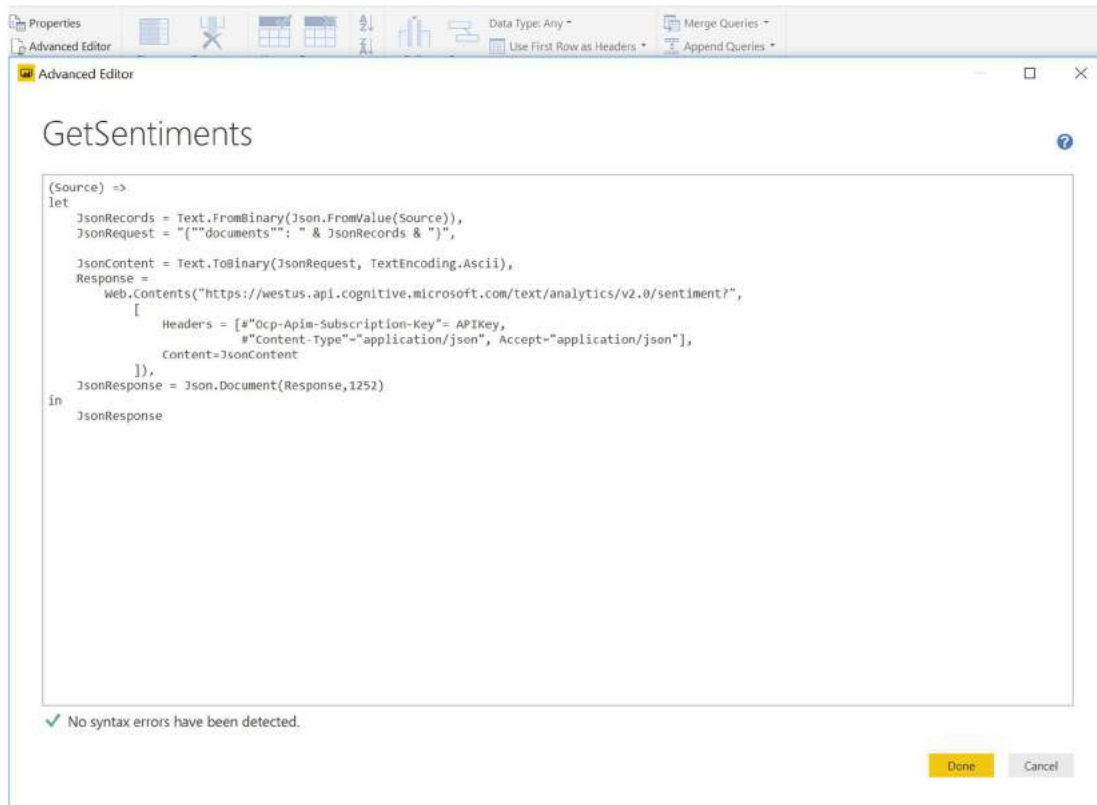
```

    JsonConvert = Text.ToBinary(JsonRequest, TextEncoding.Ascii),
    Response =
        Web.Contents("https://westus.api.cognitive.mi-
crosoft.com/text/analytics/v2.0/sentiment?",
    [
        Headers = [{"Ocp-Apim-Subscription-Key"= APIKey,
                    #"Content-Type"="application/json", Ac-
cept="application/json"}],
        Content=JsonContent
    ]),
    JsonResponse = Json.Document(Response,1252)
in
    JsonResponse

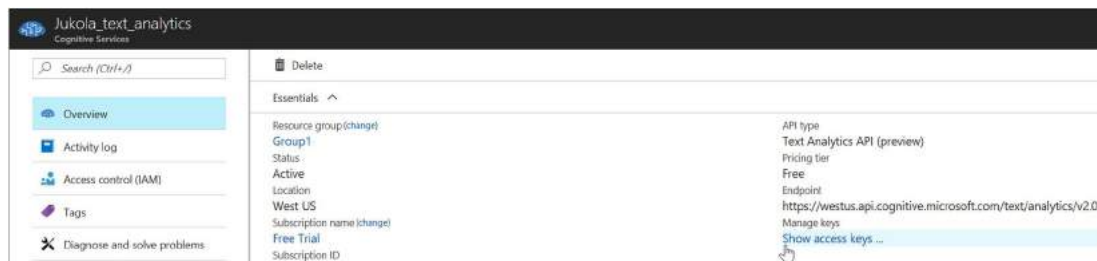
```

Scriptin arvo APIKey tulee korvata henkilökohtaisella avaimella, jonka saa *Azure*sta valitsemalla luodusta palvelusta Overview > Show access keys ... (Kuva 34). Samasta yhteydestä nähdään myös API:n osoite <https://westus.api.cognitive.microsoft.com/text/analytics/v2.0>, jota käytetään scriptissä. Scriptin luomisen jälkeen palataan takaisin edelliseen viestiin ja sen id:n sisältävään kyselyyn. Lisätään kyselyyn uusi vaihe valitsemalla Add Step (Kuva 35). Lisätään kaavan eteen juuri luomamme scriptikyselyn nimi jättäen alkuperäisen osan sulkujen sisään (Kuva 36 ja Kuva 37).

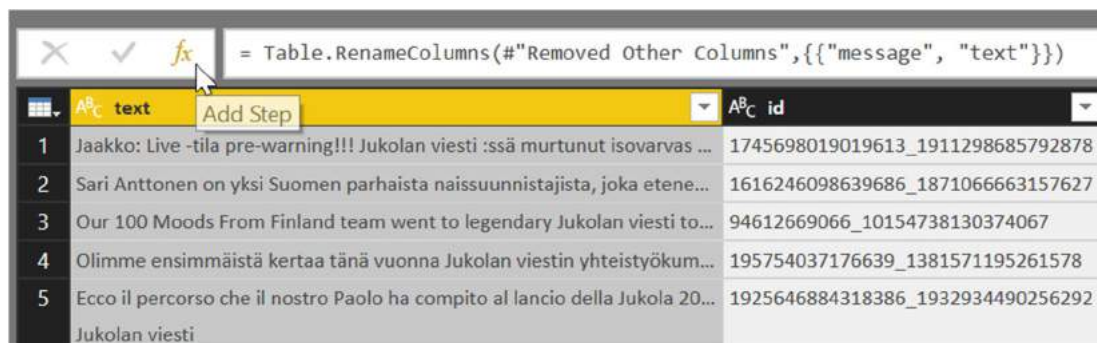
Tuloksena ovat lista documents ja lista errors. Avataan documents painamalla List linkkiä. Aukeava lista sisältää Record arvoja, eli id:n ja asenne arvon score jokaiselle viestille. Puretaan lista tauluksi valitsemalla työkalupalkista To Table (Kuva 38) ja laajennetaan taulu erillisiksi sarakkeiksi (Kuva 39). Kysely sisältää nyt jokaisen viestin tunnuksen ja sen asenteen arvona 0-1 (Kuva 40). Arvot lähellä 0 ovat negatiivisia viestejä ja arvota lähellä 1 positiivisiä. Lopuksi muutetaan score sarakkeen tyyppi valitsemalla työkalupalkista Data Type: Decimal Number. *Query editor*i voidaan sulkea ja uusi kysely painamalla yläreunan työkalupalkista Close & Apply.



Kuva 33. Scripti API:n käyttöön



Kuva 34. API avain



Kuva 35. Add Step

A screenshot of a data table interface. At the top, there is a formula bar containing the text "= #"Renamed Columns"". Below the formula bar, there are two columns: "text" and "id". The "text" column contains five rows of text, and the "id" column contains corresponding numerical IDs.

	text	id
1	Jaakko: Live -tila pre-warning!!! Jukolan viesti :ssä murtunut isovarvas ...	1745698019019613_1911298685792878
2	Sari Anttonen on yksi Suomen parhaista naissuunnistajista, joka etene...	1616246098639686_1871066663157627
3	Our 100 Moods From Finland team went to legendary Jukolan viesti to...	94612669066_10154738130374067
4	Olimme ensimmäistä kertaa tänä vuonna Jukolan viestin yhteistyökum...	195754037176639_1381571195261578
5	Ecco il percorso che il nostro Paolo ha compito al lancio della Jukola 20... Jukolan viesti	1925646884318386_1932934490256292

Kuva 36. Kaavan muokkaaminen

A screenshot of a data table interface. At the top, there is a formula bar containing the text "= GetSentiments(#"Renamed Columns")". Below the formula bar, there are two columns: "documents" and "errors". Both columns are labeled as "List".

	documents	errors
	List	List

Kuva 37. Sentimentit ja virheet listoina

A screenshot of a "To Table" dialog box. The dialog box is titled "To Table" and has a close button (X) in the top right corner. It contains the following text and controls:

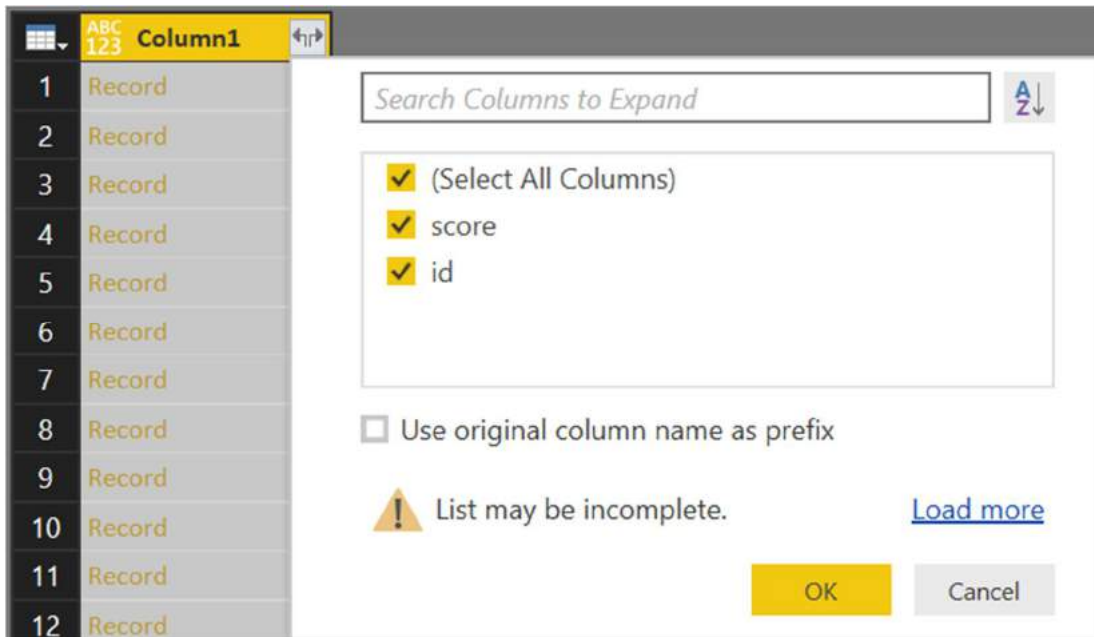
Create a table from a list of values.

Select or enter delimiter

How to handle extra columns

At the bottom right, there are two buttons: "OK" and "Cancel". A mouse cursor is pointing at the "OK" button.

Kuva 38. Lista tauluksi



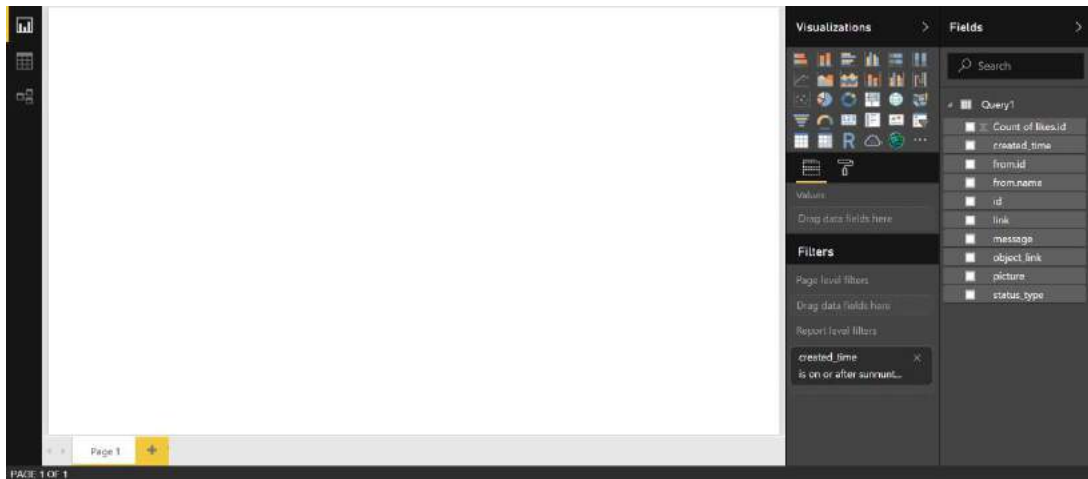
Kuva 39. Taulun laajentaminen

	ABC 123 score	ABC 123 id
1	0,5	1326015100782886_1511908572193537
2	0,5	346636812202867_658023227730889
3	0,197008398	341651939241996_1613802985360212
4	0,20767442	139395019575379_746548355526706

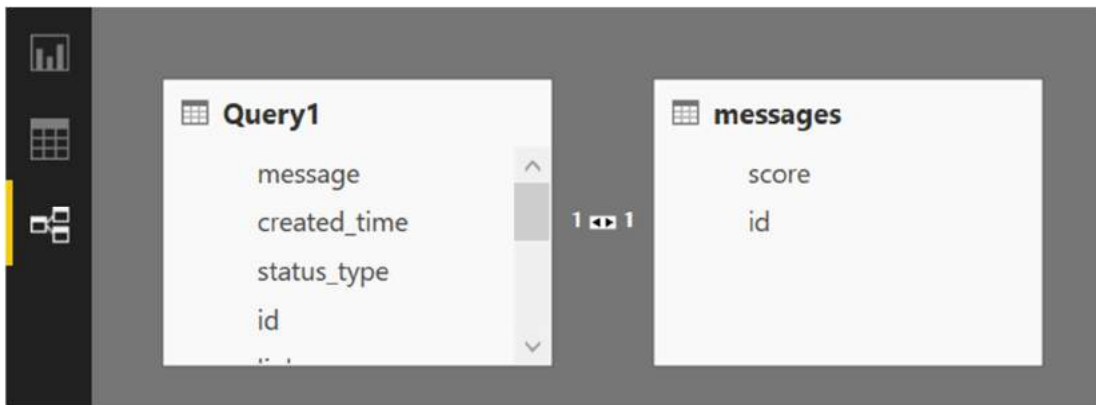
Kuva 40. Yksittäisten viestien asenteet

#### 2.4.2 Manuaalinen luominen

*Query Editorin* sulkemisen jälkeen *Power BI Desktop* aukaisee käyttöliittymän näkemysten luomiseen (Kuva 41). Kahden luomamme kyselyn välillä vallitsee automaattisesti yhteys, koska toinen kysely on kopio ensimmäisestä sisältäen saman määrän sarakkeita sekä yhteisen tekijän (Kuva 42). Kyselyjen yhteyden luomiseksi kummassakin tulee olla yhteinen tekijä, tässä tapauksessa id sarake. Yhteyden ansiosta voimme yhdistää sarakkeita kummastakin kyselystä samaan näkemykseen.

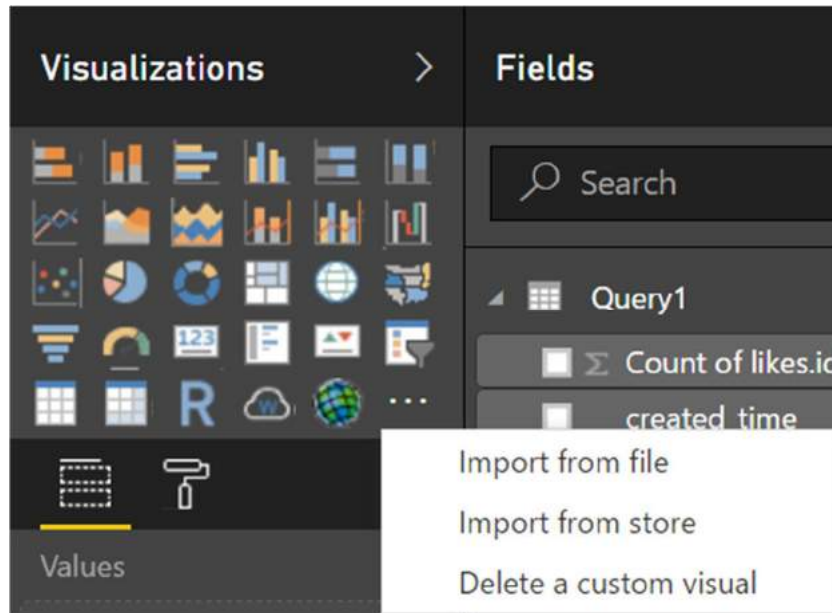


Kuva 41. Näkemysten luomisen käyttöliittymä

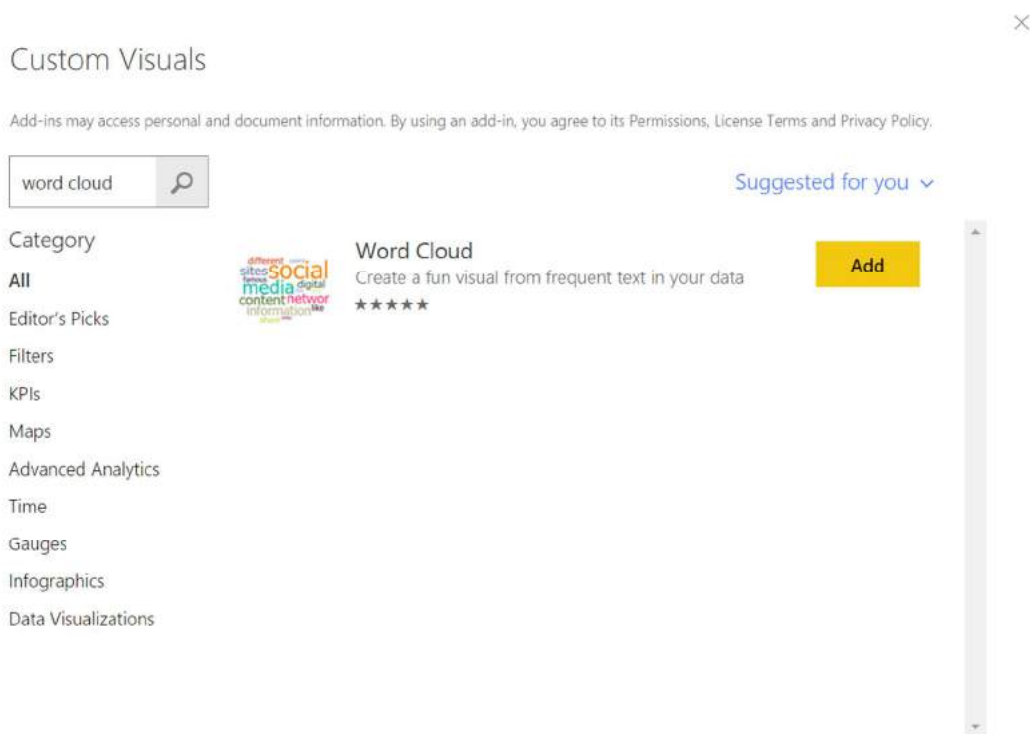


Kuva 42. Kyselyjen yhteys

Viestien sisällön visualisoimiseksi etsitään ja ladataan kaupasta sanapilvi työkalu. Avataan kauppa painamalla Visualizations valikosta . . . > Import from store (Kuva 43) ja etsitään sanapilvi hakulauseella word cloud (Kuva 44). Lisätään sanapilvi käyttöön painikkeesta Add.

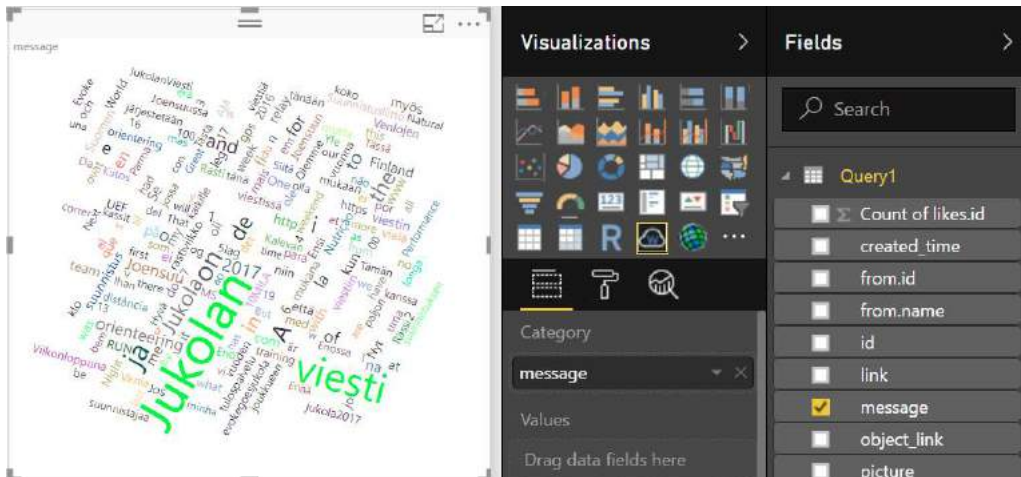


Kuva 43. Työkalukauppa



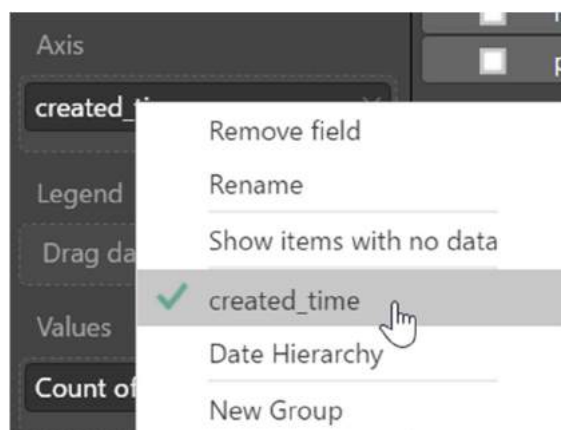
Kuva 44. Sanapilvi-työkalun lisääminen

Vastataan kysymykseen "Mikä oli viestien yleisin sisältö?" visualisoimalla viestin sisältö sanapilvellä. Valitaan visualisoinnin tyyppiä sanapilvi painamalla Visualizations valikosta word cloud. Vedetään hiiren osoittimella sarake messages sanapilven Category kenttään (Kuva 45). Sana pilvi näyttää sivun vierailijoiden viesteissään eniten käyttämät sanat fonttikoolla. Suosituimpia käytettyjä sanoja ovat muun muassa: "Julkolan", "viesti", "on", "ja", "de", "and", "for" ja "2017".



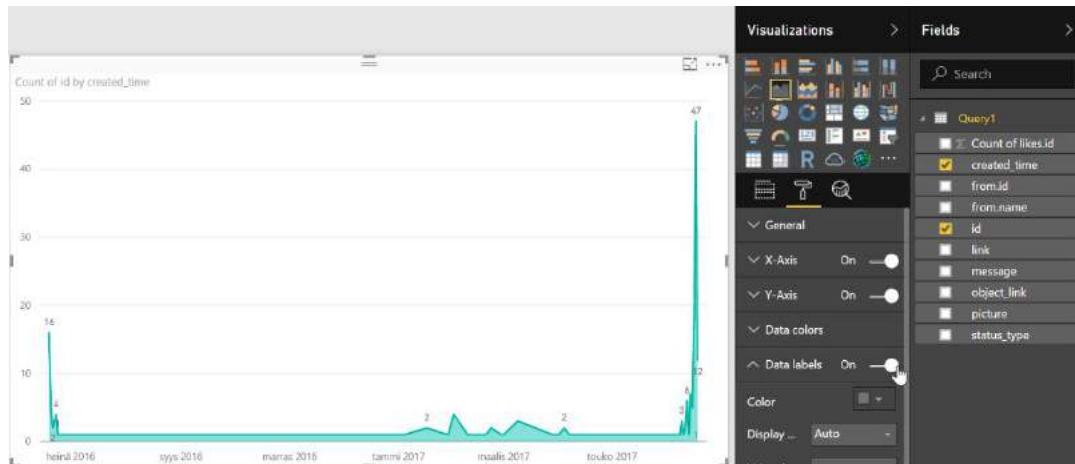
Kuva 45. Viestien yleisimmät sanat

Vastataan kysymykseen "Miten viestit jakautuivat määrällisesti tarkkailuajalle?" vertaamalla viestien päivämääriä niiden määrään. Valitaan visualisoinnin tyyppiä kaavio painamalla visualizations valikosta Area chart. Vedetään hiiren osoittimella created\_time Axis kenttään ja id Values kenttään. Date Hierarchy on automaattisesti päällä Axis kentän created\_time sarakkeessa, jonka vuoksi aika-akseli alkaa tammikuusta ja päättyy joulukuuhun. Tarkasteluväli rajoittuu kuitenkin kesäkuusta 2016 kesäkuuhun 2017, jonka vuoksi hierarkia tulee poistaa käytöstä painamalla hiiren oikealla Axis kentän created\_time valitsemalla sen sijaan järjestykseksi created\_time (Kuva 46). Kaavion arvoja selventävät luvut saadaan päälle valitsemalla Visualizations valikosta Format > Data labels. Kuvaajan mukaan käyttäjät lähettivät viestejä aktiivisimmin heti edellisen tapahtuman jälkeen sekä lähellä seuraavaa tapahtumaa (Kuva 16). Heinäkuusta 2016 toukokuuhun 2017 sivulle on lisätty käyttäjien toimesta kuukausittain vain yksittäisiä viestejä. Sivulla vierailaan lähinnä vain tapahtumakuukautena kesäkuisin.



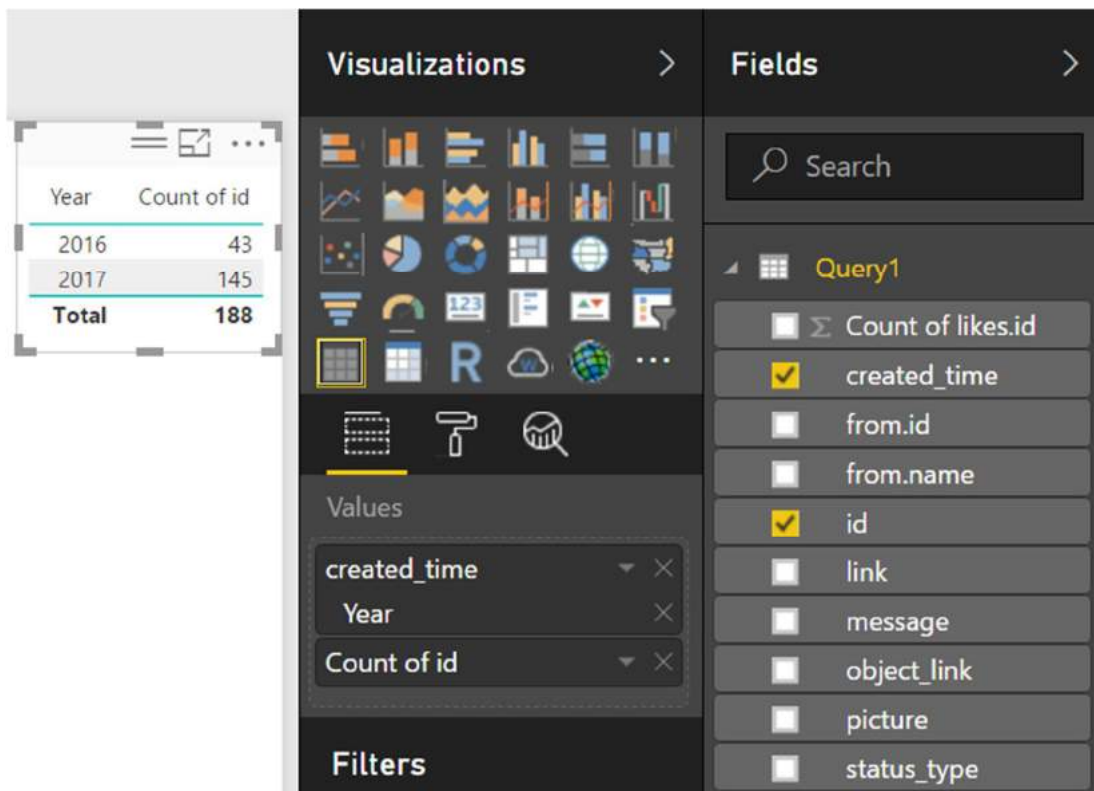
Kuva 46. Päivämäärähierarkian poistaminen käytöstä





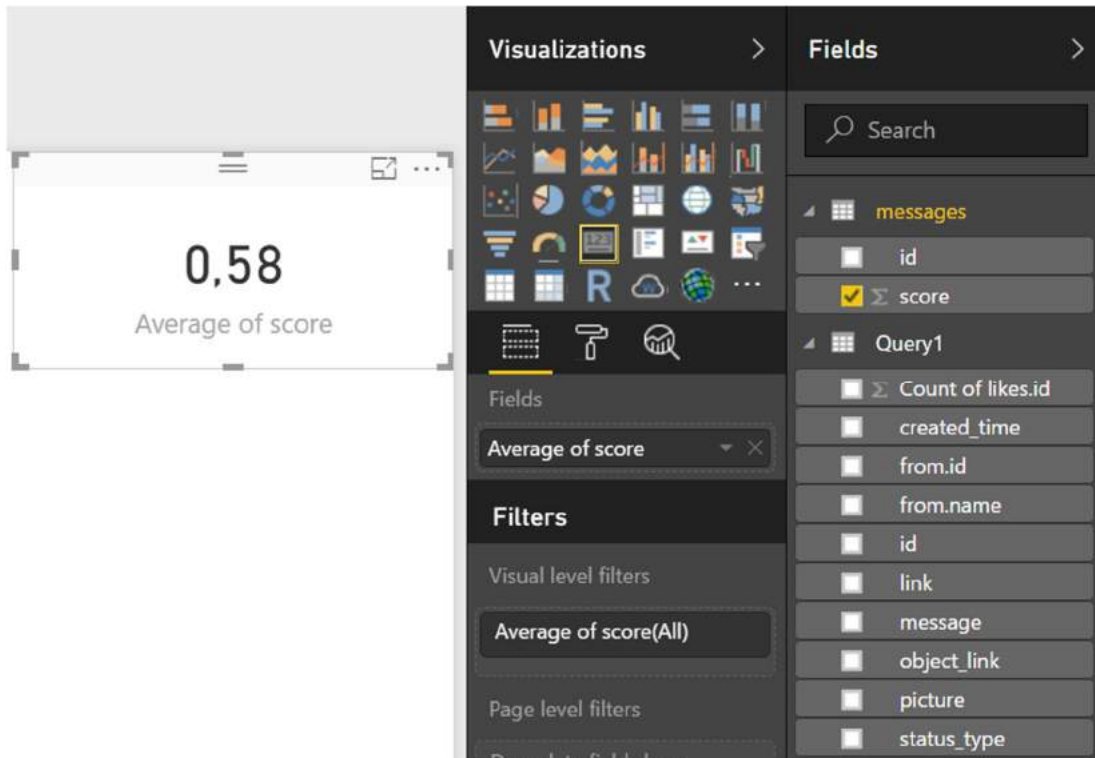
Kuva 47. Uusien viestien määrä sivulla

Vastataan kysymykseen "Kuinka paljon viestejä kirjoitettiin tarkkailuajana?" taulukoi-  
malla viestien määrät vuosittain. Valitaan visualisoinnin tyyppiksi taulu valitsemalla vi-  
sualizations valikosta Table. Vedetään hiiren osoittimella created\_time ja id  
values kenttään. Poistetaan values kentästä create\_time kohdalta Date, Month ja  
Quarter, jolloin jäljelle jää vain Year ja viestimäärät näkyvät vuosittain. Lisäksi vaih-  
detaan id tyyppiä hiiren oikealla Count, joka laskee viestien määrän. Taulukon (Kuva  
48) mukaan tarkasteluvälillä sivulle on lisätty yhteensä 188 uutta viestiä, joista suurin  
osa 145 vuoden 2017 aikana.



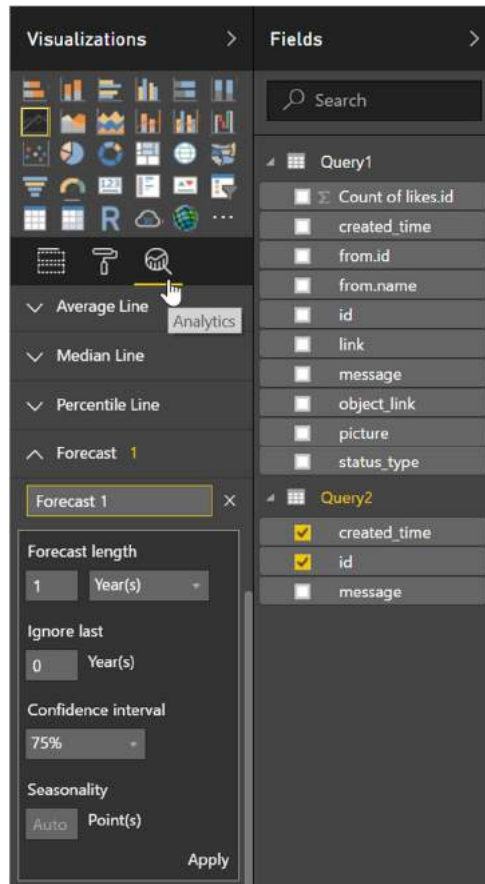
Kuva 48. Viestien yhteismäärä ja jakauma vuosittain





Kuva 50. Viestien asenteet

*Facebook* viestien määrän kehitystä voidaan ennustaa *Forecast* työkalulla, joka löytyy oikean työkalupalkin *Analytics* osiosta (Kuva 51). Sivun suhteellisen vähäisten viestimäärien ja niiden epätasaisen jakauman vuoksi ennusteita datalle ei pystytä luomaan, vaikka tarkastelu suoritettaisiin rajatun aikavälin sijaan kaikkien sivun viestien pohjalta jo vuodesta 2012 alkaen.



Kuva 51. Forecast työkalu

Valmiit näkemykset voidaan julkaista *Power BI*:n pilviversioon. Julkaisu tapahtuu painamalla yläreunan työkalupalkin painiketta `Publish` (Kuva 52). *Power BI* ilmoittaa julkaisemisen onnistumisesta ja näyttää linkin sen pilviversioon (Kuva 53). Julkaistu raportti kaikkine välisivuineen siirtyy nähtäväksi *Power BI*:n pilviversioon, jossa niitä voi käyttää ja jakaa muille. Pilviversiossa raportista voidaan myös saada pikanäkemyksiä sekä käyttää siihen *Q&A* toiminnallisuutta, joka on kuvattu luvussa 2.4.3.



Kuva 52. Publish painike

## Publishing to Power BI



✓ Success!

[Open 'jukola facebook powerbi.pbix' in Power BI](#)

[Get Quick Insights](#)



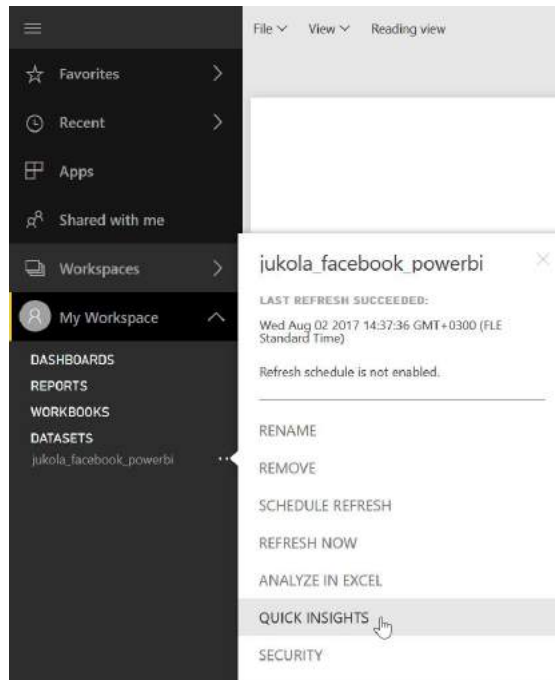
Go to the "view" tab, select "phone reports" to tailor a portrait view of your existing report for mobile viewers. [Learn more](#)

Got it

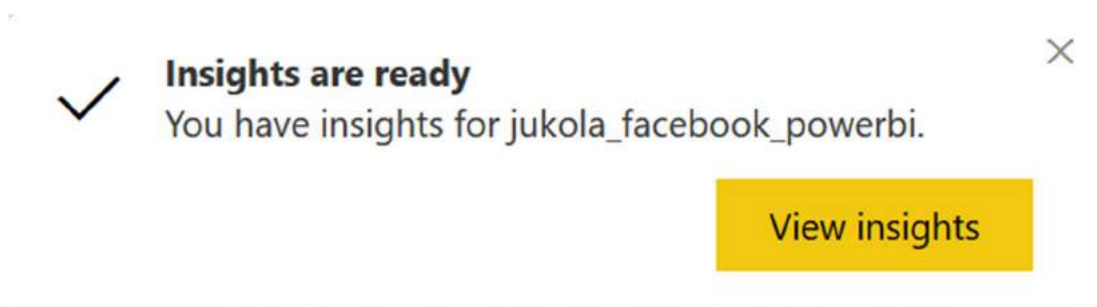
Kuva 53. Raportin julkaiseminen

### 2.4.3 Power BI pikanäkemykset ja Q&A

Julkaistu raportti näkyy pilviversio `REPORTS` osiossa tallennetulla nimellään. Haetaan pikanäkemykset painamalla datajoukkoa `DATASETS` osiossa hiiren oikealla ja valitsemalla `Quick insights` (Kuva 54). Power BI ilmoittaa löydetyistä näkemyksistä oikeaan yläkulmaan ilmestyvällä huomautuksella, josta painamalla `View insights` (Kuva 55) näkemykset avautuvat (Kuva 56). Kiinnitetään kiinnostava pikanäkemys "Count of **from.id** BY **STATUS\_TYPE**" oikean yläkulman nastasta `Pin visual` ja luodaan sitä varten uusi tilannekuva "Jukola 2017" (Kuva 57). Kiinnitetty pikanäkemys on nyt nähtävissä `DASHBOARDS` osiossa (Kuva 58). Näkemyksen perusteella pystytään vastaamaan kysymykseen "Minkä tyyppisiä viestejä lisättiin?". Kuvaajan mukaan käyttäjät ovat lisänneet eniten kuvaviestejä. 61 eri käyttäjää on lisännyt sivulle kuvia, 45 vain tekstiä ja 14 käyttäjää videoita. Lisäksi viisi viestiä on lisätty mobiilistatuksena.



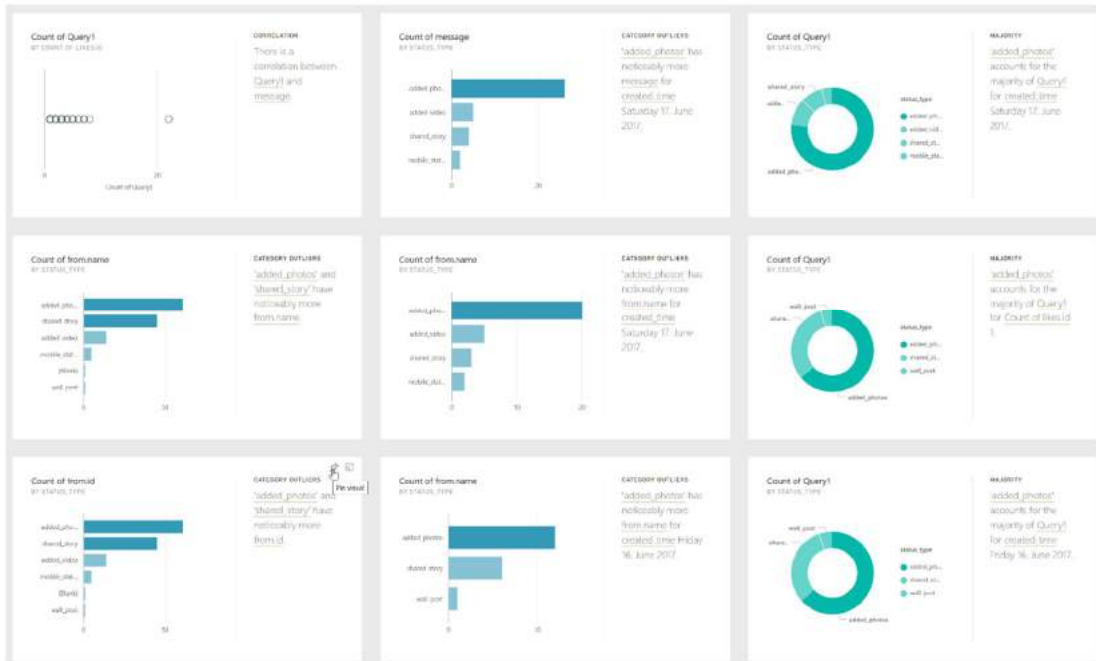
Kuva 54. Pikanäkemysten hakeminen



Kuva 55. Ilmoitus näkemysten valmistumisesta

Quick Insights for jukola\_facebook\_powerbi

A subset of your data was analyzed and the following insights were found. [Learn more](#)

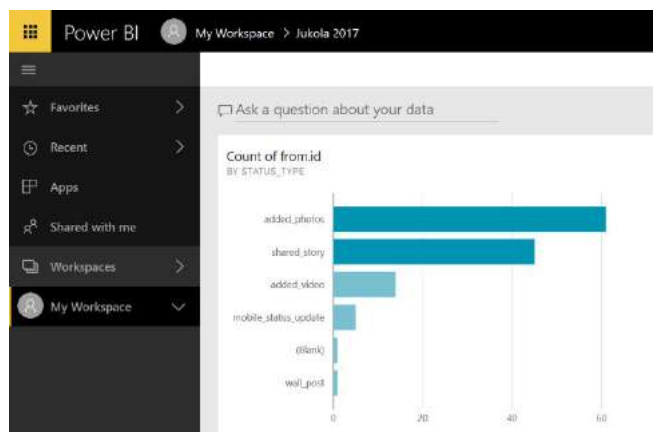


Kuva 56. Pikanäkemykset datajoukolle

The screenshot shows the 'Pin to dashboard' dialog box. On the left, a preview of the 'Count of from.id BY STATUS\_TYPE' chart is visible, showing bars for 'added...' and 'shared...'. On the right, the dialog prompts the user to select a dashboard:

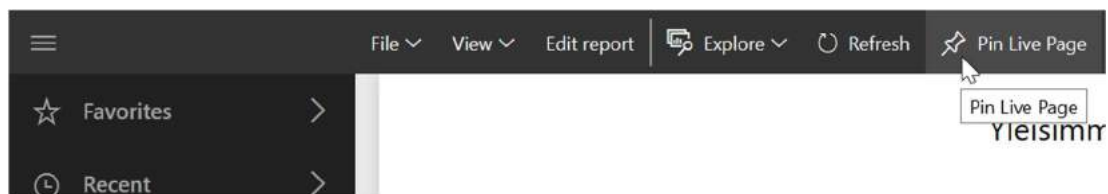
- Pin to dashboard** (with a close button 'X')
- Select an existing dashboard or create a new one.
- Where would you like to pin to?**
  - Existing dashboard
  - New dashboard
- Input field:
- Pin** (yellow button) and **Cancel** (grey button)

Kuva 57. Uuden tilannekuvan luominen



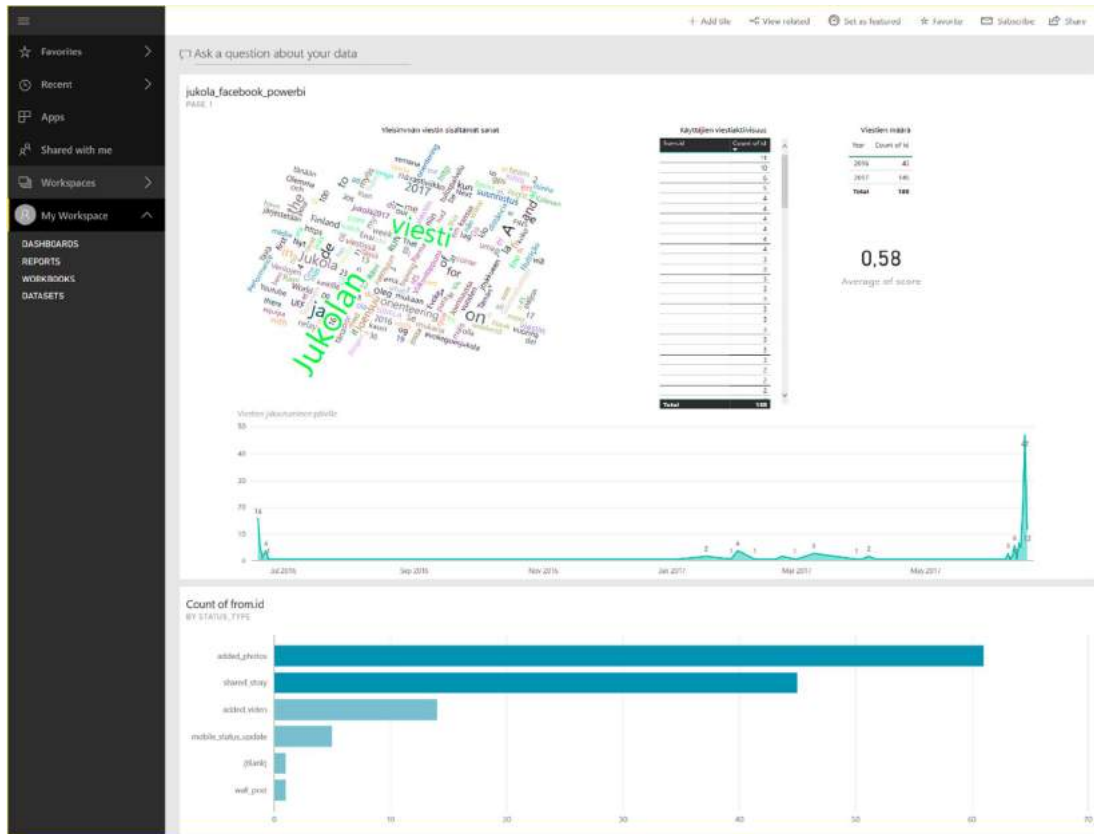
Kuva 58. Viestien tyypit

Kiinnitetään myös työpöytäversiossa luotu raportti luotuun tilannekuvaan. Aukaistaan raportti REPORTS osiossa, painetaan yläreunasta Pin Live Page (Kuva 59) ja valitaan luotu tilannekuva Jukola 2017. Raportti tulee näkyviin tilannekuvaan omana ruutunaan (Kuva 60).



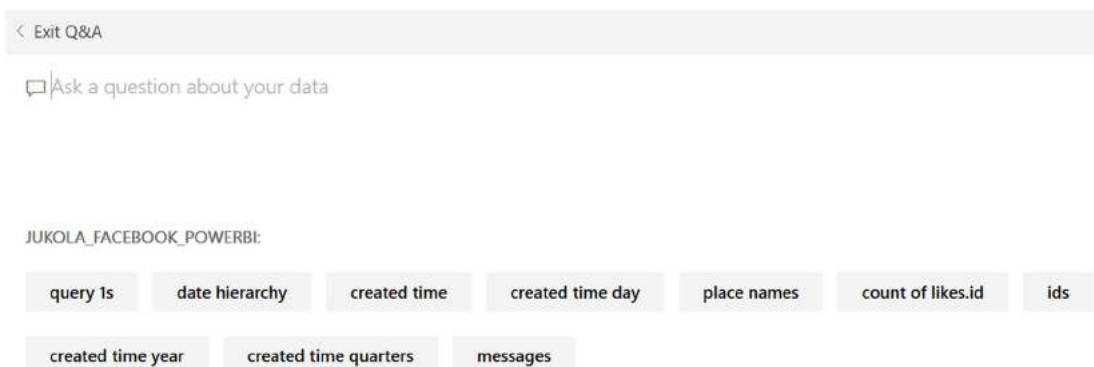
Kuva 59. Pin Live Page painike





Kuva 60. Kiinnitetty raportti (tunnisteet poistettu)

Painamalla tekstikenttää Ask a question about your data työkalu ehdottaa valmiiksi datajoukon kenttiä (Kuva 61). Vastataan kysymykseen "Kuinka moni viesti sisälsi linkkejä tai kuvia?" kysymyksillä "What is the amount of link" ja "What is the amount of picture". Kiinnitetään kysymysten tuottamat näkemykset tilannekuvaan painamalla oikeasta yläkulmasta Pin visual. Lukujen (Kuva 62) mukaan 188 viestistä 170 sisälsi ainakin yhden linkin ja 168 sisälsi ainakin yhden kuvan. Kuvan kuitenkin upotetaan viesteihin aina linkkeinä, joten vain kaksi viestiä sisälsi linkkejä, jotka eivät olleet upotettuja kuvia.

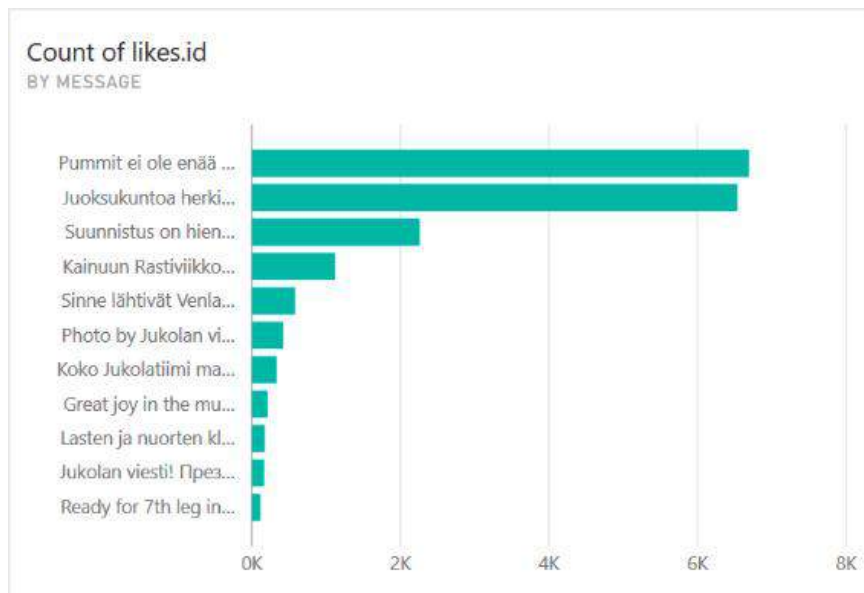


Kuva 61. Kysymyksen laatiminen

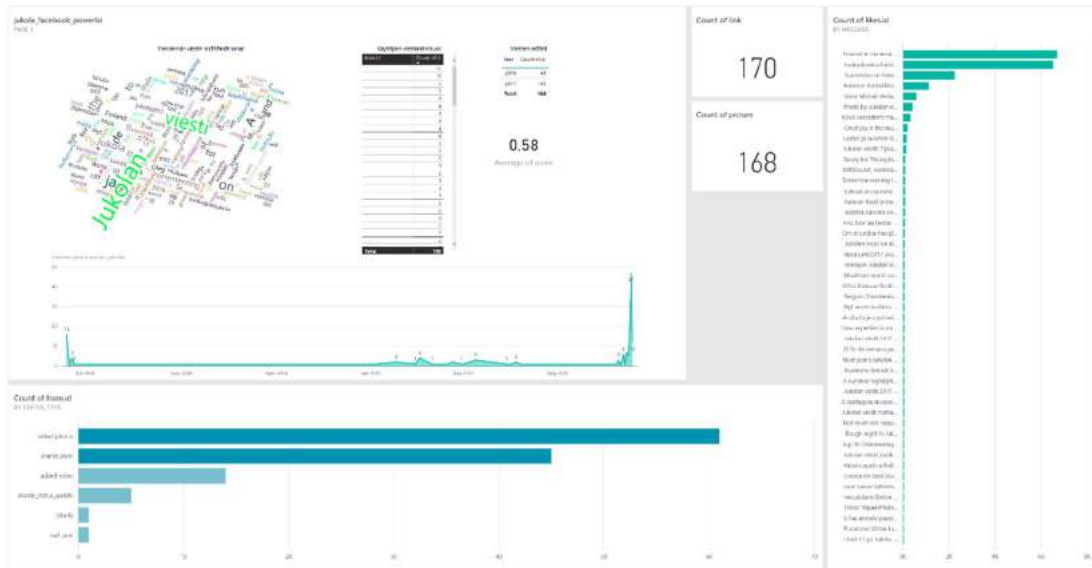


Kuva 62. Viestien sisältämät linkit ja kuvat

Vastataan kysymykseen "Mikä olivat eniten tykättyt viestit?" kysymyksellä "messages by count of likes.id". Kiinnitetään kysymyksen tuottama näkemys tilannekuvaan painamalla oikeasta yläkulmasta `pin visual`. Kuvaajan (Kuva 63) mukaan kaksi viestiä on saanut huomattavasti enemmän tykkäyksiä kuin muut viestit. Suosituimmista neljästä viestistä kolme on kirjoitettu kaksikielisesti (sekä suomeksi että englanniksi), mikä on voinut auttaa viestien leviämisessä. Tilannekuva sisältää nyt kaikki luodut näkemykset (Kuva 64).



Kuva 63. Tykättyimmät viestit



Kuva 64. Tilannekuva Facebook datasta (tunnisteet poistettu)

## 2.5 TWITTER DATA

*Power BI* ei sisällä valmista toimintoa Twitter datan hakemiseen. Tietoja voidaan hakea manuaalisesti *Query Editorissa* *Twitter API:n*<sup>18</sup> avulla samalla tavalla kuin Facebook dataa. *Twitter API* on kuitenkin huomattavasti *Facebookin API:a* suljetumpi.

*Twitter API* on jaettu julkiseen ja suljettuun. Twiittejä voidaan etsiä molempien kautta. Julkisessa *API:ssa* rajapintana toimii *Search API*<sup>19</sup>. *API:a* voidaan käyttää kyselyllä:

```
https://api.twitter.com/1.1/search/tweets.json
```

*Search API:lla* twiittejä pystytään etsimään lisäämällä kyselyn perään etsityt hakusanat. Esimerkiksi `?q=jukola,joensuu` etsii twiittejä, jotka sisältävät sanat "jukola" ja "joensuu". *API:a* käyttääkseen tulee myös tunnistautua *OAuth*<sup>20</sup> menetelmällä, joka monimutkaistaa tietojen hakemista huomattavasti. Tietoja voidaan hakea *Query Editoriin* *M scriptillä*<sup>21</sup>, joka sisältää tunnistautumisen. *Twitterin* julkinen *Search API* ei kuitenkaan indeksoi kaikkia käyttäjien lähettämiä twiittejä. Haetut twiitit löytyvät vain muutamalta edelliseltä päivältä ja nekin on rajattu enintään sataan kappaleeseen. *Search API:n* avulla voidaan seurata twiittejä lähinnä reaaliaikaisesti muutama päivä taaksepäin.

Tietojen hakeminen pidemmältä aikaväliltä edellyttää suljetun *API:n* käyttämistä. Suljetun *API:n* käyttäminen ei ole kuitenkaan mahdollista ilman maksullisia palveluita.

<sup>18</sup> <https://dev.twitter.com/overview/api>

<sup>19</sup> <https://dev.twitter.com/rest/public/search>

<sup>20</sup> <https://dev.twitter.com/oauth>

<sup>21</sup> <https://msdn.microsoft.com/en-us/library/mt211003.aspx>

Esimerkiksi *GNIP*<sup>22</sup> on *Twitterin* API alusta, joka tarjoaa maksua vastaan pääsyn suljettuun API:iin. Suljetulla API:lla tietojen hakeminen toimii samalla periaattella kuin julkisella API:lla. IBM on toteuttanut *Watson Analyticsin* *Twitter* datan hankkimisen yhteistyössä *Twitterin* kanssa, jonka ansiosta *Watson Analytics*illä tietoja voidaan hakea suljetun API:n kautta.

*Twitter* kuitenkin sallii omien twiittitietojen lataamisen CSV-tiedostoon (Kuva 65). Tiedot voi ladata kuukausittain aina viisi kuukautta taaksepäin. Tietojen avulla pystyy analysoimaan oman *Twitter* tilin aktiviteetteja, kuten tykättyjä ja jaettuja viestejä, vastaanotettuja twiittejä sekä käyttäjien painamien linkkien määrää.



Kuva 65. Käyttäjätilin twiittitietojen lataaminen

*Power BI* tarjoaa valmiita mallipohjia<sup>23</sup> *Twitter* tilin tapahtumien reaaliaikaiseen seuraamiseen. Seuranta tapahtuu Microsoft *Azure* palvelun avulla, josta tiedot ohjataan *Power BI*:hin. Myös Microsoft *Flow* sisältää *Azuren* mallipohjia<sup>24</sup> *Twitter* datan seuraamiseen, esimerkiksi uusien hakutermejä vastaavien twiittien tallentamisen Microsoft *Excel* tiedostoon.

<sup>22</sup> <https://gnip.com/>

<sup>23</sup> <https://powerbi.microsoft.com/en-us/solution-templates/brand-management-twitter/>

<sup>24</sup> <https://emea.flow.microsoft.com/en-us/templates/?category=socialMedia&sort=properties%2Fstatistics%2Finstantiations>

# 3 IBM Watson Analytics

*IBM Watson Analytics* on IBM:n täysin pilvipohjainen analytiikkatyökalu. Työkalusta on saatavilla ilmaisversio<sup>25</sup> ja ensimmäiset 30 päivää ilmaisversion yhteyteen saa *Professional* kokeiluversion, jolla pystytään esimerkiksi hakemaan suoraan dataa useista eri verkkolähteistä. *Professional* versio maksaa 80\$ kuukaudessa ja siitä hieman rajatumpi *Plus* versio 30\$ kuukaudessa. Tässä raportissa kuvatut työvaiheet on suoritettu *Watson Analytics Professional* versiolla. Yksityiskohtaiset ohjeet käyttöliittymään ja toimintoihin löytyvät sen verkkosivuilta<sup>26</sup>. IBM tarjoaa myös analytiikkatyökalua yksinomaan sosiaalisen median analysointiin - *Watson Analytics for Social Media*<sup>27</sup> kykenee hakemaan suoraan tietoja useilta eri sosiaalisen median sivustoilta ja sen kokeiluversio on käytettävissä ensimmäiset 30 päivää, mutta kokeiluversion tiedonlähteet ja toiminnallisuudet ovat hyvin rajatut. Luvussa 3.1 tarkastellaan *Watson Analytics*in toimintoja, luvussa 3.2 *Twitter* datan lataamista, luvussa 3.3 ladatun *Twitter* datan näkemyksiä, luvussa 3.4 *Facebook* datan lataamista ja luvussa 3.5 ladatun *Facebook* datan näkemyksiä.

## 3.1 WATSON ANALYTICSIN TOIMINNALLISUUDET

Tässä luvussa käydään läpi *Watson Analytics*in keskeisimmät toiminnot. Luvussa 3.1.1 esitellään datajoukkojen lataaminen ja muokkaaminen, luvussa 3.1.2 pikanäkemykset ja valmiit kysymykset, luvussa 3.1.3 näkemysten manuaalinen luominen ja luvussa 3.1.4 tilannekuvat.

### 3.1.1 Datan lataaminen ja muokkaaminen

Datajoukot *Power BI*:hin ladataan aloitusnäytön Data välilehden painikkeesta Get Data (Kuva 66). Valittavina on useita eri verkkolähteitä (*Import*), yhteyksiä (*Connection*) sekä paikallinen tiedosto (*Local file*) (Kuva 67).



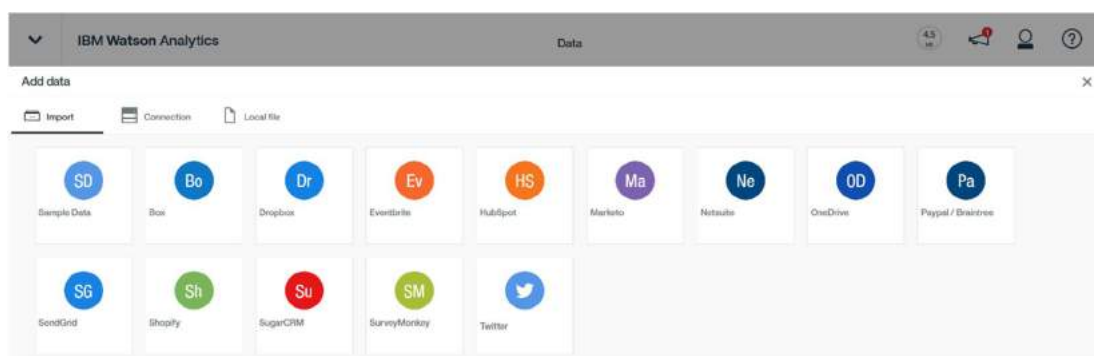
Kuva 66. Watson Analytics aloitusnäkömä

<sup>25</sup> <https://www.ibm.com/watson-analytics/pricing>

<sup>26</sup> [https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa\\_an\\_overview.2.0.0.doc/wa\\_tut\\_overview\\_for\\_docs.html](https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa_an_overview.2.0.0.doc/wa_tut_overview_for_docs.html)

<sup>27</sup> <https://www.ibm.com/us-en/marketplace/social-media-data-analysis>

Ladatut datajoukot näkyvät Data välilehdessä ja niitä muokkaamaan pääsee valitsemalla datajoukon valikosta Refine. Yksi *Watson Analytics*sin ominaisuus on datasettien laadun arvioiminen. Datajoukon laatu näkyy prosentteina sen yhteydessä ja sitä voi tarkastella lähemmin valitsemalla vasemman yläkulman Data Metrics painikkeen datajoukkoa tarkastellessa (Kuva 68). Datajoukon laatu<sup>28</sup> määräytyy datajoukon muun muassa rivimäärän, poikkeamien, puuttuvien arvojen tai muiden sisällön ongelmien perusteella. Datajoukon tarkastelun lisäksi sitä voi muokata esimerkiksi vaihtamalla sarakkeiden tyyppiä, ja lisäämällä uusia. Sarakkeita ei *Watson Analytics*issa voi poistaa, mutta ne voi piilottaa. Yksittäisten solujen arvoja ei myöskään voi muokata.



Kuva 67. Watson Analytics datalähteitä



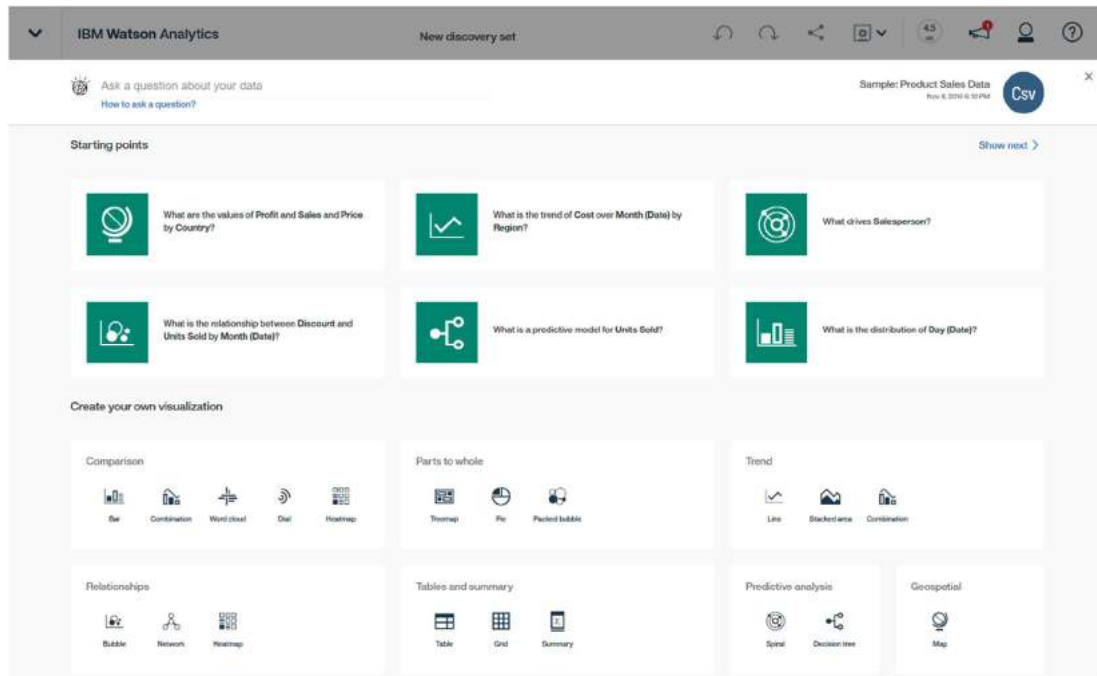
Kuva 68. Datajoukon tarkasteleminen ja sen laatu

### 3.1.2 Pikanäkemykset ja valmiit kysymykset

Näkemyksiä pääsee luomaan painamalla datajoukkoa. Tallennettuja näkemyksjoukkoja puolestaan voi selata aloitusnäytön Discover välilehdestä. *Watson Analytics* luo valmiita pikanäkemyksiä Starting points osioon. Automaattisia näkemyksiä voi parantaa muokkaamalla datajoukkoa datajoukkoa muokkaamalla. Poistamalla analysointitarkoitukselle ylimääräisiä sarakkeita, voidaan työkalua estää luomasta liikaa hyödyttömiä näkemyksiä. Lisää pikanäkemyksiä voi nähdä klikkaamalla Show next

<sup>28</sup> [https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa\\_an\\_overview.2.0.0.doc/wa\\_an\\_hlp\\_data\\_qual.html](https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa_an_overview.2.0.0.doc/wa_an_hlp_data_qual.html)

painiketta `Starting points` osion oikeassa yläkulmassa. Visualisoinnin tyyppi näkyy logona kysymyksen vasemmalla puolella. Näkemyksen voi aukaista painamalla sitä.



Kuva 69. Näkemysten luomisen näkymä

Erilaisia kysymystyyppejä on 18 kappaletta ja ne voit nähdä `How to ask a question?` painikkeesta sivun yläreunassa. Kysymystyypit jaetaan viiteen kategoriaan: vertailu (`compare data`), suhteet ja mallit (`understand relationships and identify patterns`), kokoaminen (`aggregate data`), järjestä ja suodata (`sort and filter`) sekä ennustus (`predict data`) (Kuva 70). Kysymyksen voi kirjoittaa kokonaan itse tai käyttää valmiita kysymyspohjia drop-down valikoilla. Kysymyksen esittämisen jälkeen näkemystä ei visualisoida suoraan vaan *Watson Analytics* ehdottaa erilaisia näkemyksiä kysymyksen pohjalta `Starting points` kohdassa. Kysymyksen parhaiten sopivat (`Most relevant`) näkemykset ovat listalla ensimmäisenä ja osittain sopivat (`Somewhat relevant`) listan lopussa.



Select a category

- All available examples
- Variety pack
- Compare data
- Understand relationships and identify patterns
- Aggregate data
- Sort and filter data
- Predict data
- All available examples

How do the values of  compare by  ?

How do the values of  compare by  and  ?

What is the breakdown of  by  ?

What is the breakdown of  by  and  ?

by

What is the relationship between  and  by  ?

What is the contribution of  over  by  ?

What is the average  per  ?

What is the total  per  ?

What is the maximum  by  ?

What is the minimum  by  ?

What are the top  by  ?

What are the top   by  ?

What are the bottom  by  ?

What are the bottom   by  ?

What is the count of  by  ?

What drives  ?

Kuva 70. Kysymystyytit

### 3.1.3 Näkemysten luominen

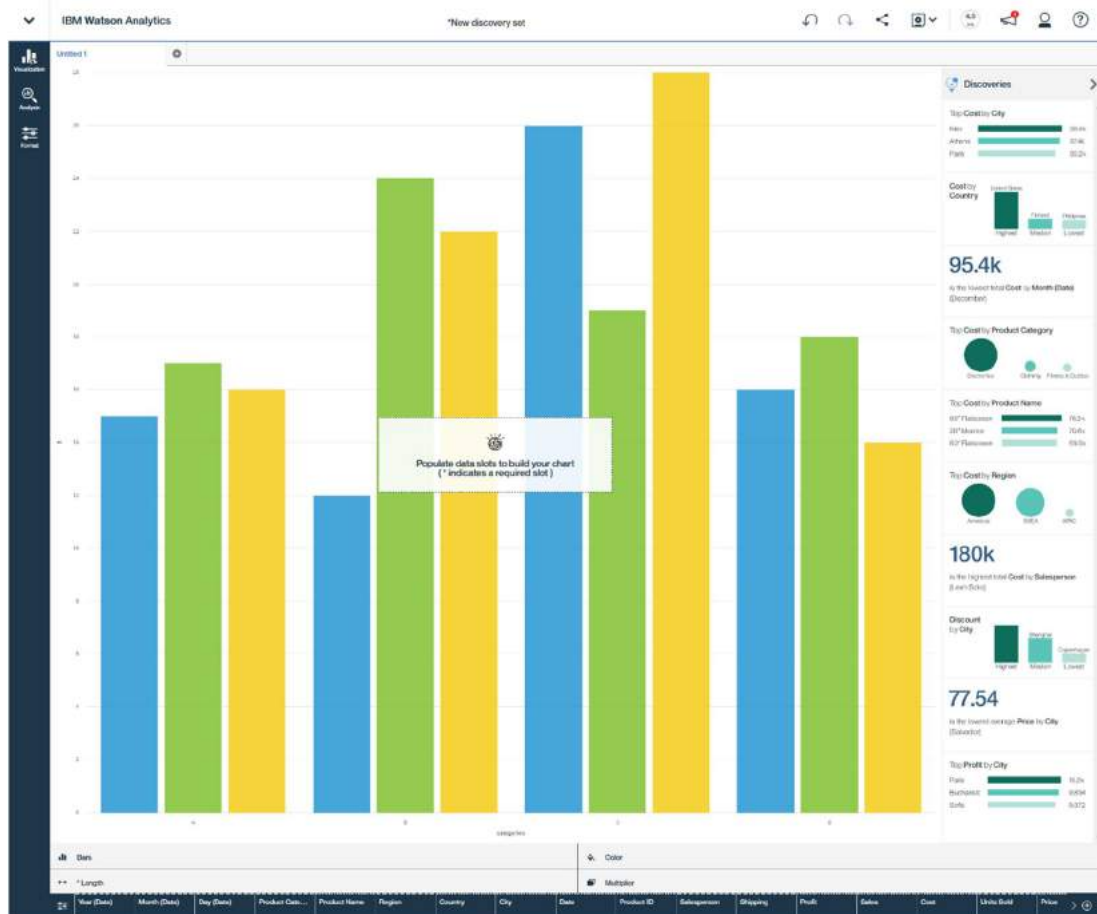
Visualisointeja voi luoda manuaalisesti alhaalta *Create your own visualization* osiosta. Visualisointi voidaan valita 7:stä eri luokasta: vertailu (*Comparison*), osia kokonaisuudesta (*Parts to whole*), trendi (*Trend*), suhteet (*Relationships*), taulut ja yhteenvedot (*Tables and summary*), ennustava analyysi (*Predictive analysis*) ja sijainti (*Geospatial*). Valitaan tyypiksi pylväsdiagrammi (*Bar*) klikkaamalla sitä, joka aukaisee visualisointinäkömän (Kuva 71). Alhaalla ovat visualisointiin



sisällöllisesti vaikuttavat kentät, joihin muuttujia voi vetää hiiren osoittimella alareunasta.

Visualisointinäkylässä pystytään muodostamaan ja muokkaamaan näkemyksiä sisällöllisesti ja visuaalisesti. Visualisoinnin tyyppiä voi vaihtaa milloin tahansa vasemman reunan *Visualization* painikkeesta. *Analysis* painikkeesta voi suorittaa lisäanalyyskejä luodun näkemyksen pohjalta ja *Format* painikkeesta muokata visualisoinnin ulkonäköä. Oikean reunan *Discoveries* osiossa on esillä datajoukon näkemyksiä, joita työkalu on löytänyt vertaamalla näkemyksessä käytettyjä datatyppejä muihin tyyppeihin.

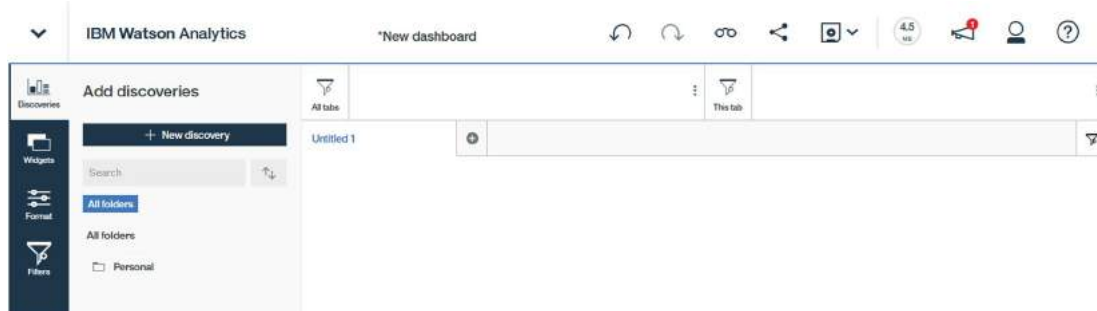
Näkemyksen voi nimetä uudelleen valitsemalla sen yläreunan välilehdestä *Rename*. Näkemykset tallennetaan klikkaamalla oikean yläkulman *Save* kuvaketta ja se on ehdonäköisesti myöhemmää tilannekuvan luomista varten. Näkemyksiä voidaan poistaa yläreunasta sulkemalla näkemyksen välilehti ja uusia näkemyksiä luoda lisäämällä uusi välilehti, joka avautuu kysymystenluontinäkymäksi. Vähintään yksi välilehti tulee olla avoin, joten työkalu ei anna poistaa ensimmäistä välilehteä ennen kuin toinen on avattu.



Kuva 71. Näkemysten luominen

### 3.1.4 Tilannekuva

Tilannekuva luodaan aloitusnäytön *Display* välilehdestä painamalla *New display*. Tilannekuvan sijaan voidaan luoda infograafi (eli valitaan *Dashboardin* sijaan *Infographic*), joka on yksi pitkä ja kapea sivu vieritettävä sivu. Tilannekuvaan voidaan lisätä vapaasti näkemyksiä vetämällä niitä vasemmasta sivupalkista *Add discoveries* osiosta. Näkemyksiä voi siirrellä ja niiden kokoa muuttaa hiiren osoittimella vetämällä. Näkemykset ovat myös vuorovaikutuksessa keskenään eli tarkastelua voi kohdentaa valitsemalla esimerkiksi vuosiluvun näkemyksestä, jolloin myös muut näkemykset muuttuvat esittämään valintaa. Valmis tilannekuva voidaan lähettää sähköpostiin, ladata, twiitata tai linkittää painamalla yläreunan *Share* painiketta.



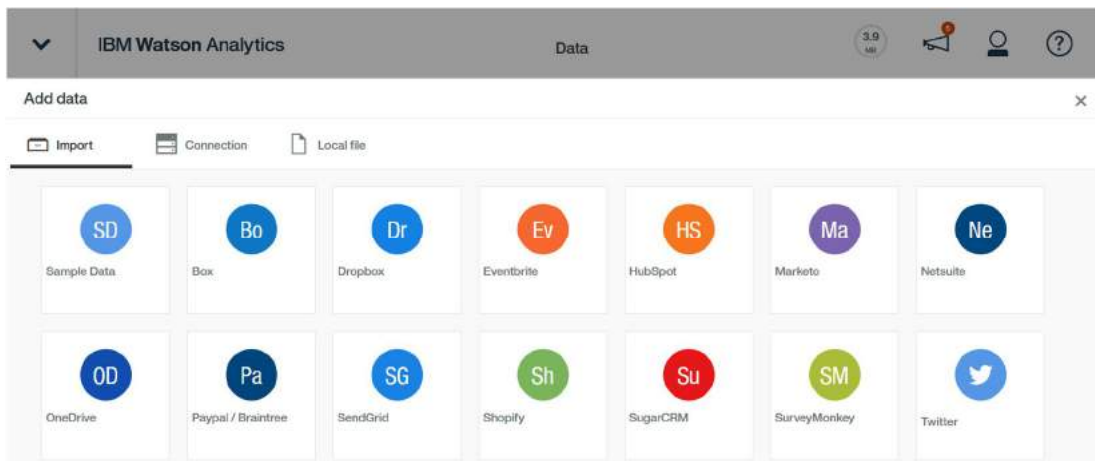
Kuva 72. Tilannekuva näkymä

## 3.2 TWITTER DATAN HANKKIMINEN

*Watson Analytics* tukee suoraan *Twitteriä* tiedonlähteenä. Avataan lähteet painamalla *Data* osiosta *New Data* (Kuva 73) ja valitaan lähteeksi *Twitter* (Kuva 74).



Kuva 73. Watson Analytics aloitusnäky ja New data painike



Kuva 74. Tiedonlähteen valinta

Tiedonlähteen valinnan jälkeen määritellään haluttu sisältö haettavaksi (Kuva 75). *Twitter* dataa haetaan hashtagien avulla. Jukolan viestin viralliset hashtagit ovat #jukola ja #joensuu. Koska tapahtumalla on useampi hashtag, on tärkeää valita vain twiitit, jotka sisältävät kummankin hashtagin `include all of the hashtags` valinnalla, jotta voidaan välttyä hakemasta tapahtumaan kuulumattomia twiittejä.

#### Select the Twitter data you want

Enter up to 10 hashtags separated by spaces, for example: #ibmWatson #analytics

#jukola #joensuu

Include any of the hashtags  Include all of the hashtags

Language

All languages

Enter dates and times in your local time. Your current time zone is UTC+-3 hours.

Start date	Time	End date	Time
2016-06-19	00:00	2017-06-18	23:59

Data asset name

#jukola #joensuu

Show estimates ^

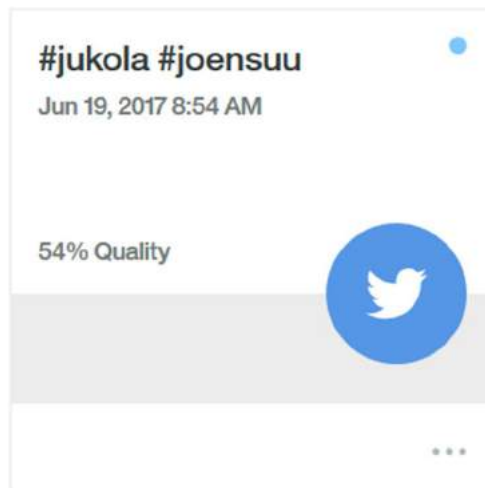
Tweets available: 125      Size (MB): < 0,1

Kuva 75. Twitter sisällön valitseminen

Haun alkamispäiväksi voidaan asettaa edellisen vuoden 2016 tapahtuman päättymisajankohdasta 2017 tapahtuman päättymispäivään. Arvioitu asetettujen kriteerien löytämien twiittien määrä näkyy `Show estimates` osiossa alhaalla. *Watson Analytics* löy-

tää hashtageilla 125 twiittiä. Suuri osa käyttäjistä ei tarkkailujaksolla käyttänyt tapahtuman verkkosivuilla ilmoittamia virallisia hashtageja, *Watson Analytics* löytää tarkkailualajalta esimerkiksi hashtagilla #jukolanviesti 77 twiittiä ja hashtagilla #jukola 55 viestiä. Lisäksi osa viesteistä, jotka on suunnattu suoraan käyttäjälle @JukolanViesti, eivät välttämättä sisällä erikseen Jukolaan liittyviä hashtageja. Eri hakujen tuottamia datajoukkoja pystytään yhdistämään<sup>29</sup>. Esimerkiksi kahden eri hashtag yhdistelmän perusteella haetut datajoukot voidaan yhdistää yhdeksi. Tässä dokumentissa analysoimme virallisten hashtagien twiittejä. Twiitit haetaan oikean alakulman `Import` painikkeella.

Haettu datasetti tulee näkyviin työkalun etusivulle (Kuva 76). Haetun Jukolan aineiston laatu on 54 %. Datajoukon laatua ja sisältöä voidaan tarkastella tarkemmin painamalla datajoukon valikosta `... > Refine`.



Kuva 76. Ladattu datajoukko

Datan laatua laskee erityisesti puuttuvat arvot (Kuva 77). Tietoja esimerkiksi twiittaajan sijainnista ei voida hakea, jos käyttäjä ei ole näitä tietoja profiiliinsa lisännyt. Tämän vuoksi sarakkeissa suurin osa kyseisistä arvoista on tuntemattomia (unknown). Myös viestien asennetiedot puuttuvat suurimmasta osasta, koska asenteita ei voida etsiä suomenkielisestä tekstistä. Huomattavaa on myös, että *Twitter* ei sisällä mahdollisuutta ilmoittaa profiilissa sukupuoltaan, vaan tämä määräytyy algoritmipohjaisesti, jonka vuoksi myös suurin osa näistä arvoista ovat tuntemattomia tai sukupuoli saattaa olla jopa väärä. Tämä aiheuttaa arvojen epätasaisen jakautumisen, joka laskee laatua. Laatu kertoo siitä, kuinka hyvin data sopii päätöksentekoon, mutta myös alhaisen laadun datasta voidaan löytää hyödyllisiä näkemyksiä.

---

<sup>29</sup> [https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa\\_an\\_overview.2.0.0.doc/wa\\_an\\_join\\_data\\_sets.html](https://www.ibm.com/support/knowledgecenter/en/SS4QC9/com.ibm.solutions.wa_an_overview.2.0.0.doc/wa_an_join_data_sets.html)



Kuva 77. Datajoukon sarakkeiden laatu

Twiittien sisältämä tieto nähdään taulukkomuodossa (Kuva 78). Haetut tiedot ovat aika (Posted time), maa (Author country), osavaltio (Author state), kaupunki (Author city), hashtagit (Hashtags), vastaus (In reply to), kieli (Language), hakuehtoja vastaavat hashtagit (Matching Hashtags), lähetetty (Posted from), asenne (Sentiment), asenteen positiiviset merkit (Sentiment positive signals), asenteen negatiiviset merkit (Sentiment negative signals), twiitin tyyppi (Tweet type), lähettäjän nimi (Author name), twiitin id (Tweet), lähettäjän sukupuoli (Author gender), lähettäjän aikavyöhyke (Author time zone), lähettäjän tykättyjen viestien määrä (Author favorite count), lähettäjän ystävämäärä (Author friend count), lähettäjän seuraajien määrä (Author follower count), lähettäjän listattujen twiittien määrä (Author listed count) twiitin jakokerrat (Retweet count) ja lähettäjän twiittien määrä (Author tweet count). Siirrytään näkemyksien luomiseen palaamalla aloitusnäkymän Data välilehteen ja painamalla datajoukkoa tai painamalla `Create discovery set` (Kuva 79) aloitusnäkymän `Discover` välilehdestä.

IBM Watson Analytics #jukola #joensuu

Year (Posted from)	Month (Posted from)	Day (Posted from)	Hour (Posted from)	Author country	Author state	Author city	#joensuu	#jukola	Hashtags	In reply to
2017	April	13	16	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	June	18	19	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	May	29	19	finland	eastern finla...	kuopio	1	1	["#joensuu", "...	No
2017	January	19	10	finland	eastern finla...	joensuu	1	1	["#joensuu", "...	No
2017	June	17	5	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	April	27	6	finland	southern finl...	helsinki	1	1	["#joensuu", "...	No
2017	February	2	19	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	April	3	15	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	June	18	7	finland	unknown	unknown	1	1	["#joensuu", "...	No
2017	February	2	8	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	June	17	16	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2016	October	28	12	unknown	unknown	unknown	1	1	["#maavimat...	No
2016	July	18	11	sweden	västra götaland	göteborg	1	1	["#joensuu", "...	No
2017	June	18	5	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2016	December	1	15	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	January	25	14	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2016	June	30	8	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	January	18	16	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2016	June	29	16	finland	southern finl...	helsinki	1	1	["#joensuu", "...	No
2017	June	15	5	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	March	9	9	finland	province of w...	valkeakoski	1	1	["#suomi100", "...	No
2017	January	18	15	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	May	28	2	finland	eastern finla...	joensuu	1	1	["#jukola", "#jo...	No
2016	December	30	17	unknown	unknown	unknown	1	1	["#joensuu", "...	No
2017	June	18	9	finland	eastern finla...	joensuu	1	1	["#fingerpori", "...	No
2017	January	1	11	unknown	unknown	unknown	1	1	["#suomi100", "...	No
2016	December	1	7	unknown	unknown	unknown	1	1	["#joensuu", "...	No

Kuva 78. Hashtageilla löydetyt twiitit



Kuva 79. New discovery set painike

### 3.3 NÄKEMYKSET TWITTER DATASTA

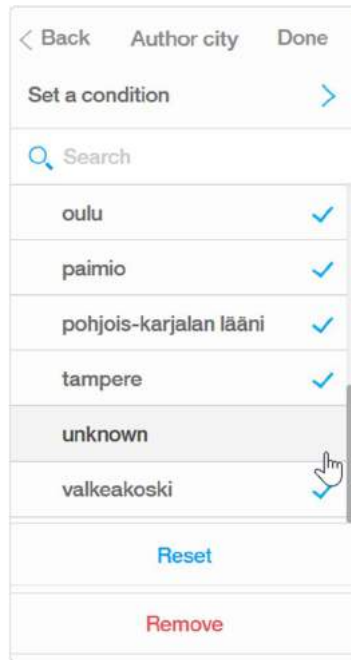
Tässä luvussa kuvataan näkemysten luominen *Watson Analytics*issä *Twitter* datalle. Luvussa 3.3.1 luodaan näkemyksiä manuaalisesti ja luvussa 3.3.2 pikanäkemysten ja kysymysten avulla.

#### 3.3.1 Manuaalinen luominen

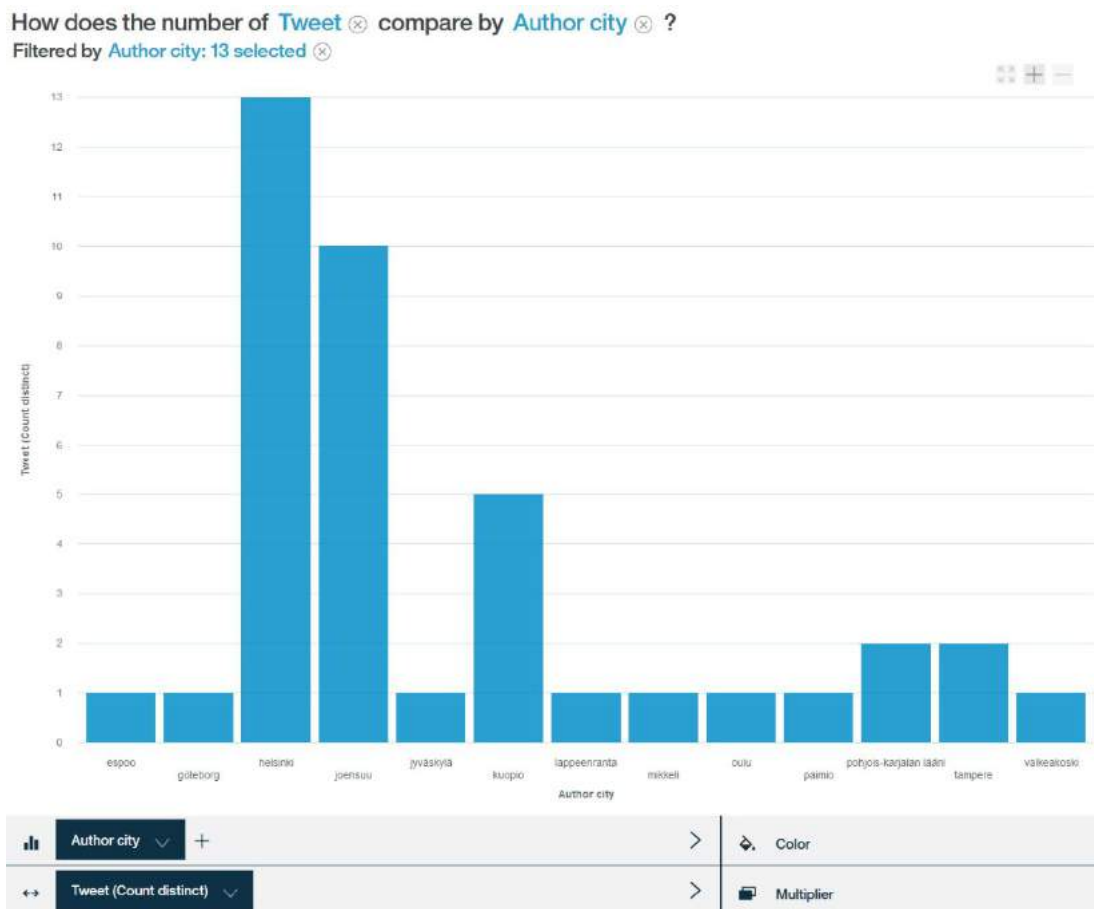
Visualisointityypit näkyvät *Create your own visualization* osiossa (Kuva 80). Vastataan kysymykseen "*Mikä oli käyttäjien sijainti?*" vertaamalla käyttäjien kaupunkoja twiittien lukumääriin. Valitaan visualisoinnin tyyppiksi pylväsdiagrammi painamalla *Create your own visualization* osiosta *Bar*. Vedetään hiiren osoittimella *Author city Bars* kenttään ja *Tweet Length* kenttään. Koska osa kaupungeista ei ole tiedossa, tuloksissa on paljon *unknown* arvoja. Painetaan *Bars* kentän *Author city > Set a condition* ja otetaan valinta pois *unknown* arvosta (Kuva 81). Näin diagrammi ei luo omaa pylvästä tuntemattomille arvoille. Samalla tavalla voidaan poistaa myös valinta *All*, joka loisi oman pylvään yhteenvetona kaikkien twiittien määrästä. Kuvaajan (Kuva 82) mukaan *helsinki* on ollut aktiivisin kaupunki *Twitter*issä kolmelta twiitillä ja seuraavana on joensuu kymmenellä twiitillä. Avataan uusi välilehti seuraavalle näkemykselle yläreunan *Add a new tab* painikkeesta.



Kuva 80. Näkemysten manuaalinen luominen

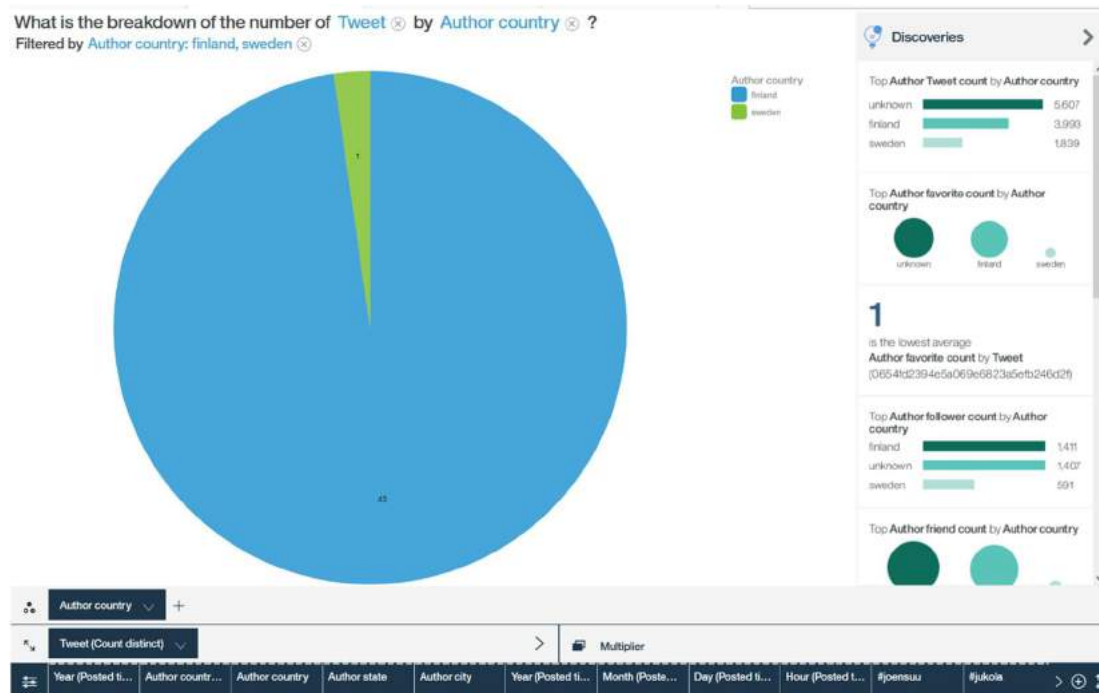


Kuva 81. Tulosten suodattaminen



Kuva 82. Käyttäjien kaupungit

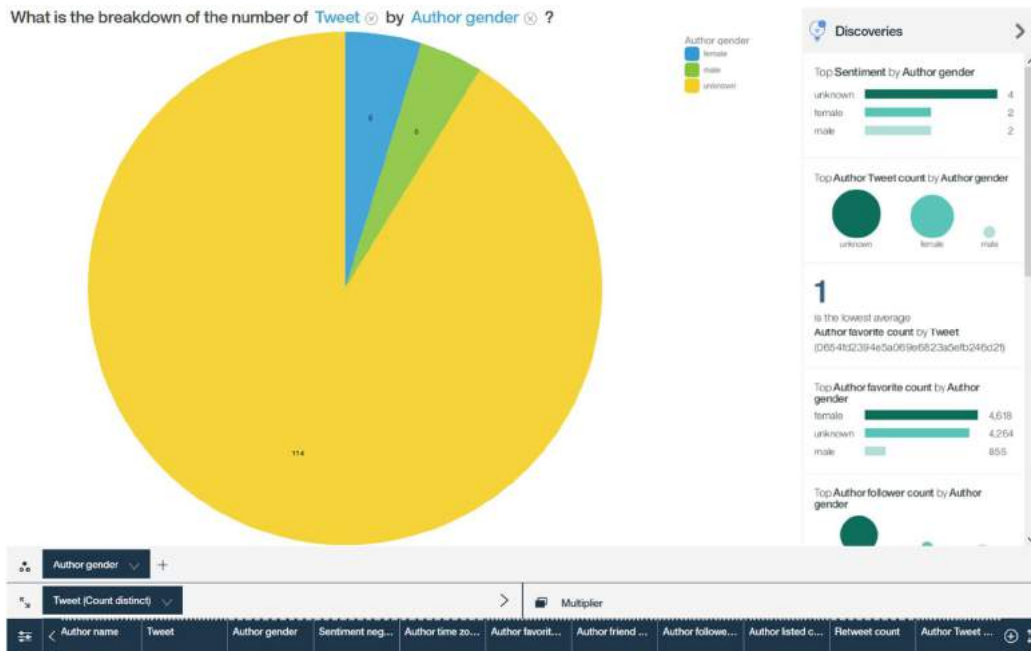
Selvitetään myös käyttäjien jakautuminen valtioittain. Valitaan visualisoinnin tyyppiä piirakkadiagrammi painamalla *Create your own visualization* osiosta *Pie*. Vedetään hiiren osoittimella *Author country Segments* kenttään ja *Tweet Size* kenttään. Kuvaajan (Kuva 83) mukaan suurin osa twiiteistä on kirjoitettu suomalaisten käyttäjien toimesta ja vain yksi twiitti on ruotsalaisen kirjoittama. Yhteensä 44 twiittiä sisältää sijainnin. Avataan uusi välilehti seuraavalle näkemykselle yläreunan *Add a new tab* painikkeesta.



Kuva 83. Käyttäjien sijainti valtioittain

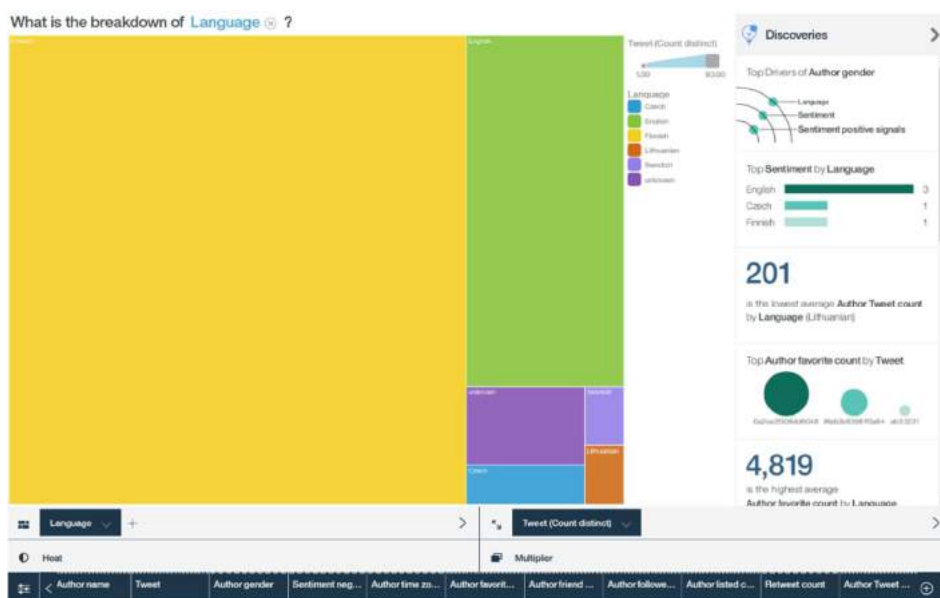
Vastataan kysymykseen "Mikä oli käyttäjien sukupuoli jakauma?" vertaamalla käyttäjien sukupuolia twiittien määrään. Valitaan visualisoinnin tyyppiä piirakkadiagrammi painamalla *Create your own visualizations* osiosta *Pie*. Vedetään hiiren oikealla *Author gender Segments* kenttään ja *Tweet size* kenttään. Kuvaajan (Kuva 84) mukaan vain 6 käyttäjää voidaan todeta naisiksi ja 5 miehiksi. Tällä tuloksella ei pystytä tekemään luotettavia johtopäätöksiä käyttäjien sukupuolen jakautumisesta. Avataan uusi välilehti seuraavalle näkemykselle yläreunan *Add a new tab* painikkeesta.





Kuva 84. Käyttäjien sukupuolijakauma

Vastataan kysymykseen "Millä kielillä twiittejä kirjoitettiin?" vertaamalla viestien kieliä niiden määrään. Valitaan visualisoinnin tyypiksi aluekartta painamalla Create your own visualizations osiosta [Areamap](#). Vedetään hiiren osoittimella Language Area hierarchy kenttään ja Tweet size kenttään. Kuvaajan (Kuva 85) mukaan suurin osa twiiteistä on kirjoitettu suomeksi (Finnish) ja seuraavaksi suosituin kieli on englanti (English). Muita yksittäisiä kieliä ovat ruotsi (Swedish), liettua (Lithuanian) ja tšekki (Czech). Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.



Kuva 85. Twiittien kielet

### 3.3.2 Pikanäkemykset ja valmiit kysymykset

Pikanäkemykset näkyvät `Starting points` osiossa (Kuva 86). Valitaan kiinnostava pikanäkemyks `What are the most common values of Day (Posted time)?`, joka on kuvakkeen mukaisesti tyypiltään sanapilvi. Pikanäkemyks vastaa kysymykseen *"Miten twiitit jakautuivat määrällisesti tarkkailuajalle?"*. Valittu pikanäkemyks aukeaa tarkasteltavaksi (Kuva 87). Kysymyksen mukaisesti sanapilvi esittää twiittien määrää päivämäärinä. Sanapilven mukaan eniten Jukolaan liittyviä twiittejä on lähetetty tapahtuman aikana 18 kesäkuuta 2017. Näkemyksen luomisessa käytetyt kentät eli aika `Year (Posted time)` sanoina ja twiittien määrä `Author Tweet count` fonttikokona näkyy alhaalla. Kenttiä voi lisätä tai vaihtaa alhaalla samalla tavalla kuin näkemyksen manuaalisessa luomisessa. Huomattavaa on, että *Watson Analytics* valitsi käytettäväksi kentäksi `Author tweet count`, joka tarkoittaa käyttäjän kaikkia lähettämiä viestejä. Näkemyks ei siis kuvasta täysin etsittävää vastausta. Lisäksi sanapilvi ei ole selkeä tapa visualisoimaan päivämääriä. Painetaan `Visualization` ja vaihdetaan tyyppiä viivadiagrammi `Line`, sekä muutetaan alhaalta `Author tweet count > Tweet` sillä haluamme tietää tapahtumaan liittyvien twiittien määrän emmekä käyttäjän kokonaismäärä. Kysymykseksi muodostuu `What is the trend of the number of Tweet over Day (Posted time)?` ja näkemyksistä visualisoidaan nyt viivadiagrammilla (Kuva 88). Kuvaajan mukaan twiittejä on kirjoitettu huomattavasti enemmän tapahtuman aikana. Avataan uusi välilehti seuraavalle näkemykselle yläreunan `Add a new tab` painikkeesta.



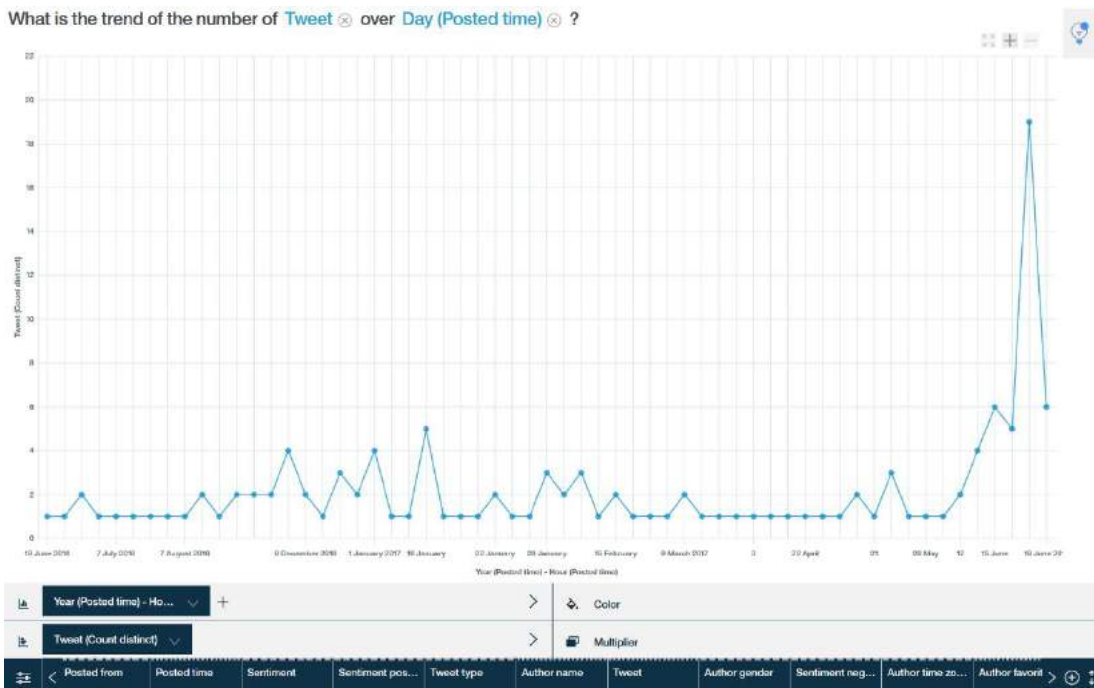
Kuva 86. Datajoukon pikanäkemykset

What are the most relevant values of **Day (Posted time)** for **Author Tweet count** ?



Kuva 87. Pikanäkemyksen tarkastelu

What is the trend of the number of **Tweet** over **Day (Posted time)** ?



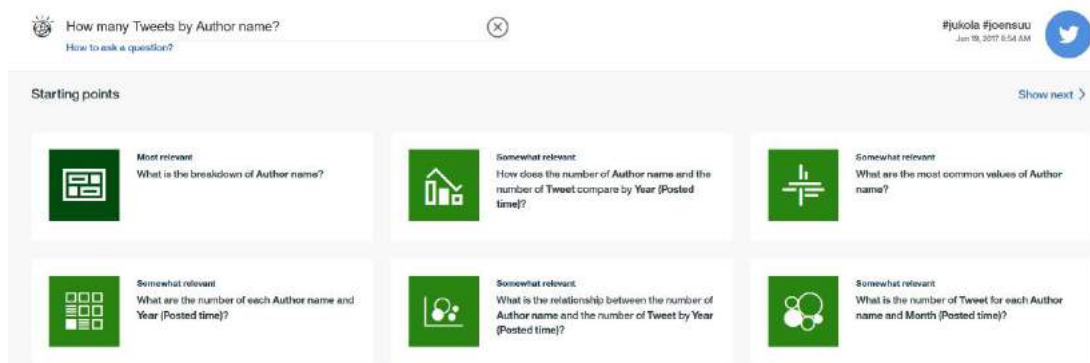
Kuva 88. Käyttäjien lähettämien twiittien määrä tarkkailuajalla

Kysymykset kirjoitetaan yläreunan Ask a question about your data kenttään (Kuva 89). Vastataan kysymykseen "Montako twiittia yksittäiset henkilöt lähettivät?" kysymyksellä "How many Tweets by Author name?" Painetaan ensimmäistä näkemystä what is the breakdown of **Author name**?, joka aukaisee sen tarkastelua

ja muokkaamista varten (Kuva 90). Kuvaajan (Kuva 91) mukaan yksi käyttäjä on lähettänyt huomattavasti muita enemmän twiittejä. Avataan uusi välilehti seuraavalle näkemykselle yläreunan `Add a new tab` painikkeesta.



Kuva 89. Kysymyskenttä

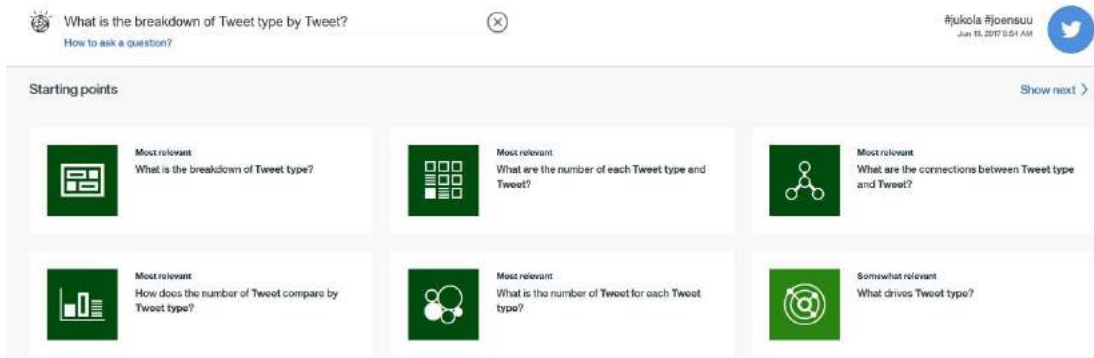


Kuva 90. Kysymyksen pohjalta ehdotetut näkemykset

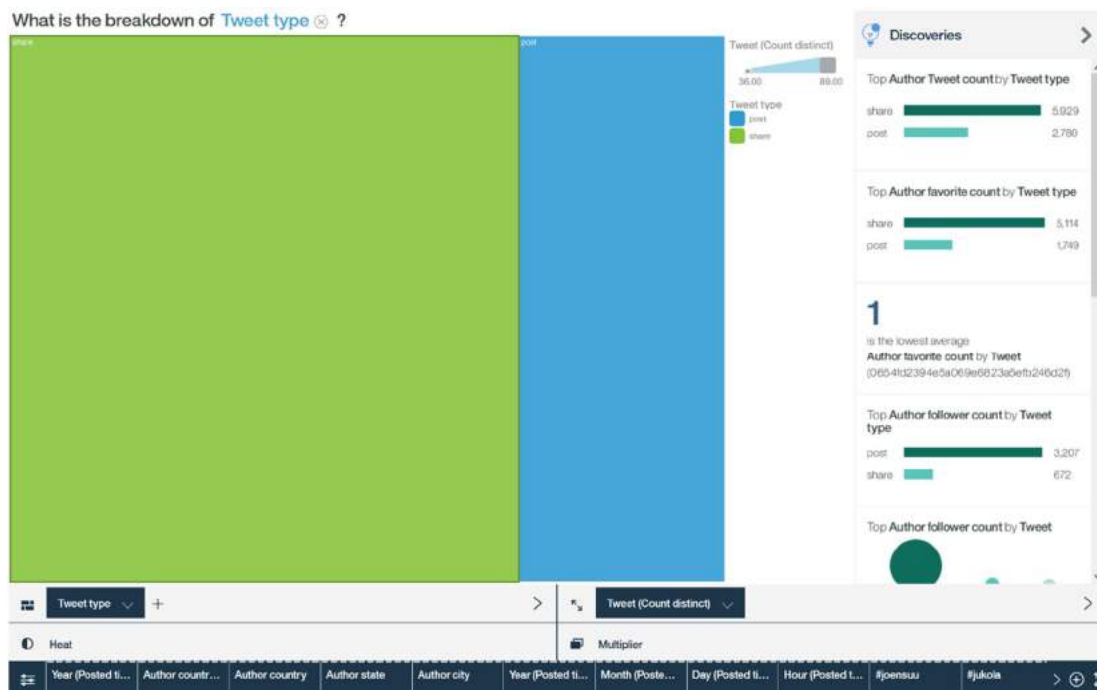


Kuva 91. Twiittien määrä lähettäjäkohtaisesti (tunnisteet poistettu)

Vastataan kysymykseen "Minkä tyyppisiä twiittejä lähetettiin?" kysymyksellä "What is the breakdown of Tweet type by Tweet?" ja valitaan ensimmäinen näkemys what is the breakdown of **Tweet type**? (Kuva 92). Kuvaajan (Kuva 93) mukaan suurin osa twiiteistä ei ole käyttäjien itsensä kirjoittamia vaan muiden luomia jaettuja twiittejä. Vain 38 twiittiä on käyttäjien itsensä kirjoittamia. Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.

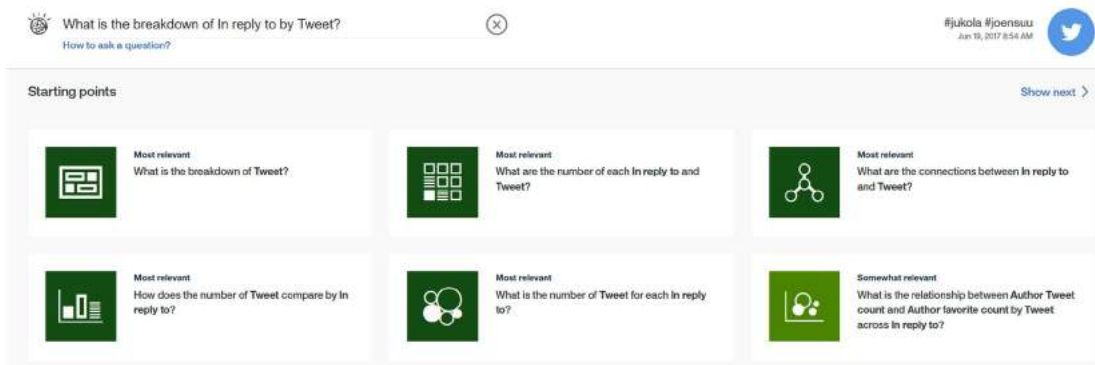


Kuva 92. Kysymyksen pohjalta ehdotetut näkemykset

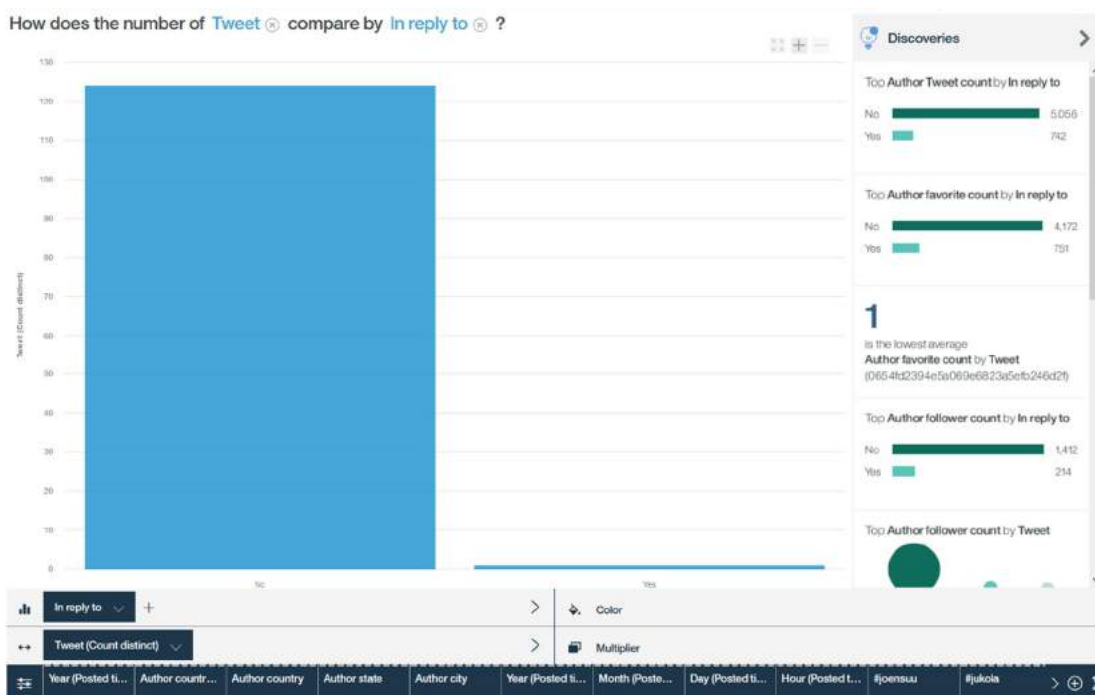


Kuva 93. Twiittien tyyppien jakautuminen

Vastataan kysymykseen "Kuinka suuri osa twiiteistä oli vastauksia?" kysymyksellä "What is the breakdown of In reply to by Tweet?" ja valitaan neljäs näkemys How does the number of **Tweet** compare by **In reply to**? (Kuva 94). Kuvaajan (Kuva 95) mukaan vain yksi twiitti on vastaus jonkin toisen käyttäjän twiittiin. Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.



Kuva 94. Kysymyksen pohjalta ehdotetut näkemykset



Kuva 95. Vastausten määrä twiiteistä

Vastataan kysymykseen "Mitkä olivat jaetuimmat twiitit?" kysymyksellä "What is the Retweet count for each Tweet?" ja valitaan ensimmäinen näkemys what is the breakdown of **Tweet**? (Kuva 96). Kuvaaja (Kuva 97) näyttää twiittien tunnukset ja niiden saamat jakokerrat. Selvitetään myös twiitin lähettäneen käyttäjänimi lisäämällä Area hierarchy kohtaan ensimmäiseksi Author name (Kuva 98). Suosituin twiitti on jaettu 30 kertaa, kaksi seuraavaksi suosituinta viestiä ovat saaneet 30 ja jakokertaa ja neljänneksi suosituin 25 jakokertaa. Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.

What is the Retweet count for each Tweet? How to ask a question? #jukola #joensuu Jun 19, 2017 8:54 AM

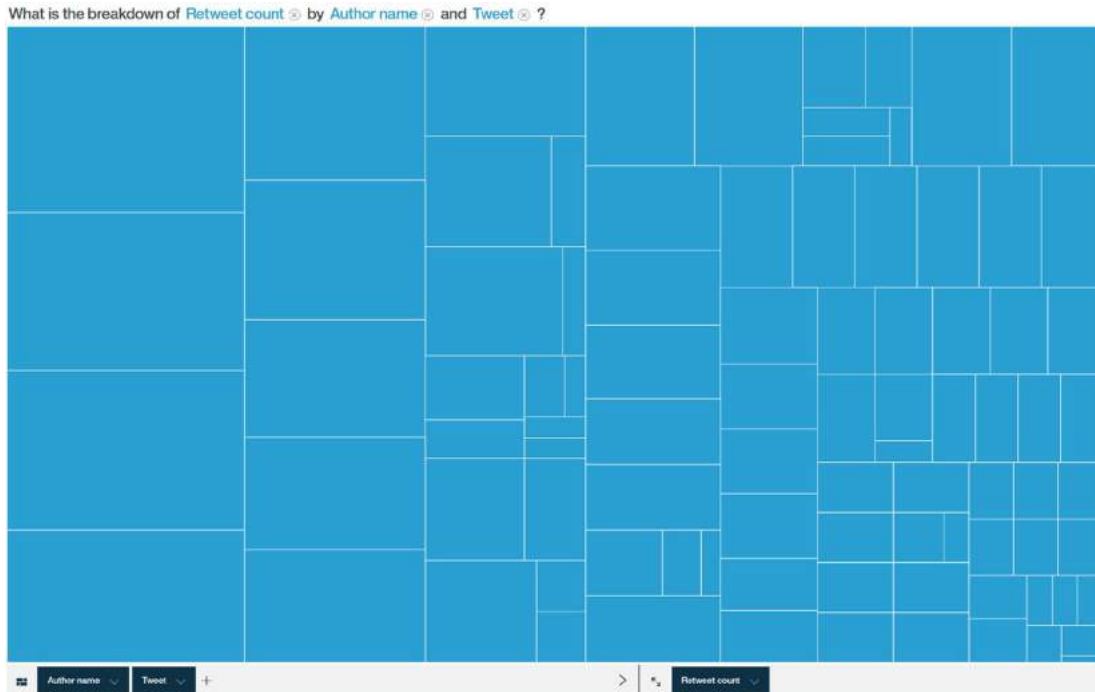
**Starting points** Show next >

- What is the breakdown of Author follower count by Author gender and Language?
- How do the values of Author friend count compare by Author gender and Author country?
- How do the values of Author Tweet count compare by Retweet count and Sentiment negative signals?
- What is the relationship between Author listed count and Author favorite count by Author gender?
- What are the most common values of Day (Posted time)?
- What are the number of each Year (Posted time) and Author country?

Kuva 96. Kysymyksen pohjalta ehdotetut näkemykset

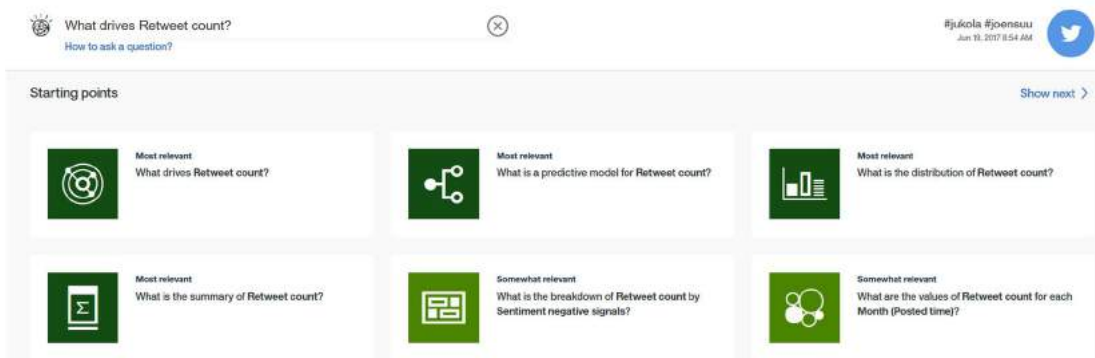


Kuva 97. Jaetuimmat twiitit



Kuva 98. Jaetuimmat twiitit käyttäjänimillä (tunnisteet poistettu)

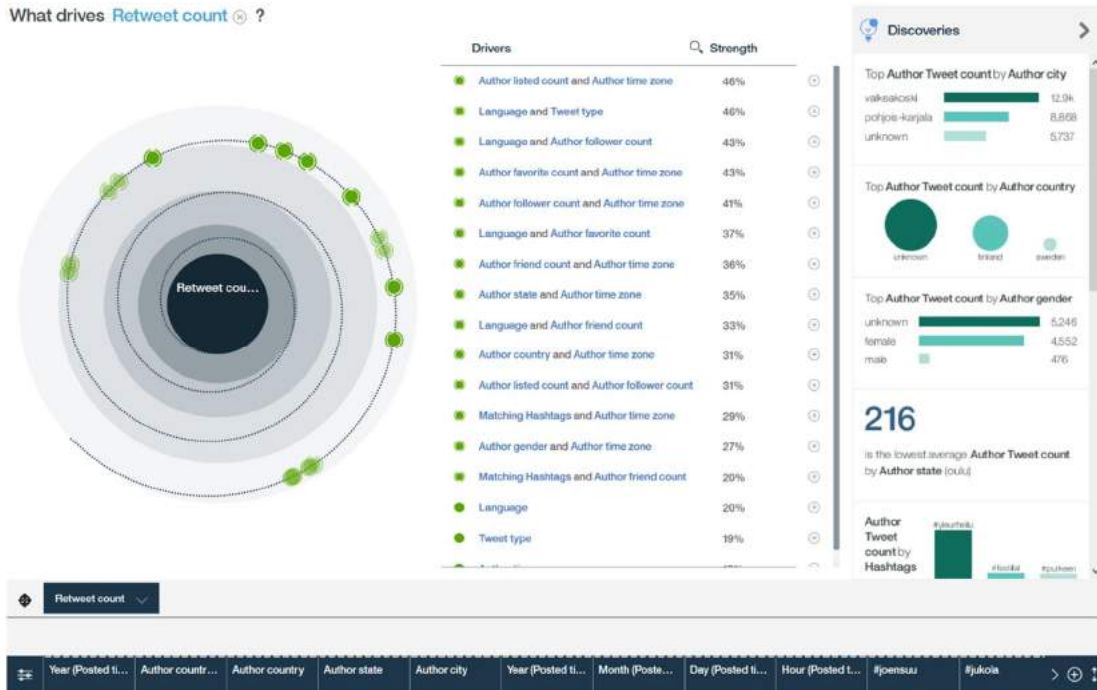
Vastataan kysymykseen "Mitkä tekijät vaikuttavat twiittien suosioon?" kysymyksellä "What drives Retweet count?" ja valitaan ensimmäinen näkemys What drives **Retweet** count? (Kuva 99). Kuvaaja (Kuva 100) näyttää oikeassa laidassa twiittien jakokertoihin vaikuttavat muuttujat. Listan kärjessä ovat twiitin kirjoittajan listattujen twiittien määrä ja aikavyöhyke sekä kieli ja twiitin tyyppi. Painamalla plussaa listan oikealla puolella on mahdollista nähdä yksityiskohtaisempaa tietoa ajureista. Esimerkiksi painamalla plussaa **Language** and **Tweet type** kohdasta selviää, että suosituimpia twiittejä ovat suomeksi kirjoitetut jaetut twiitit (Kuva 101). Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.



Kuva 99. Kysymyksen pohjalta ehdotetut näkemykset

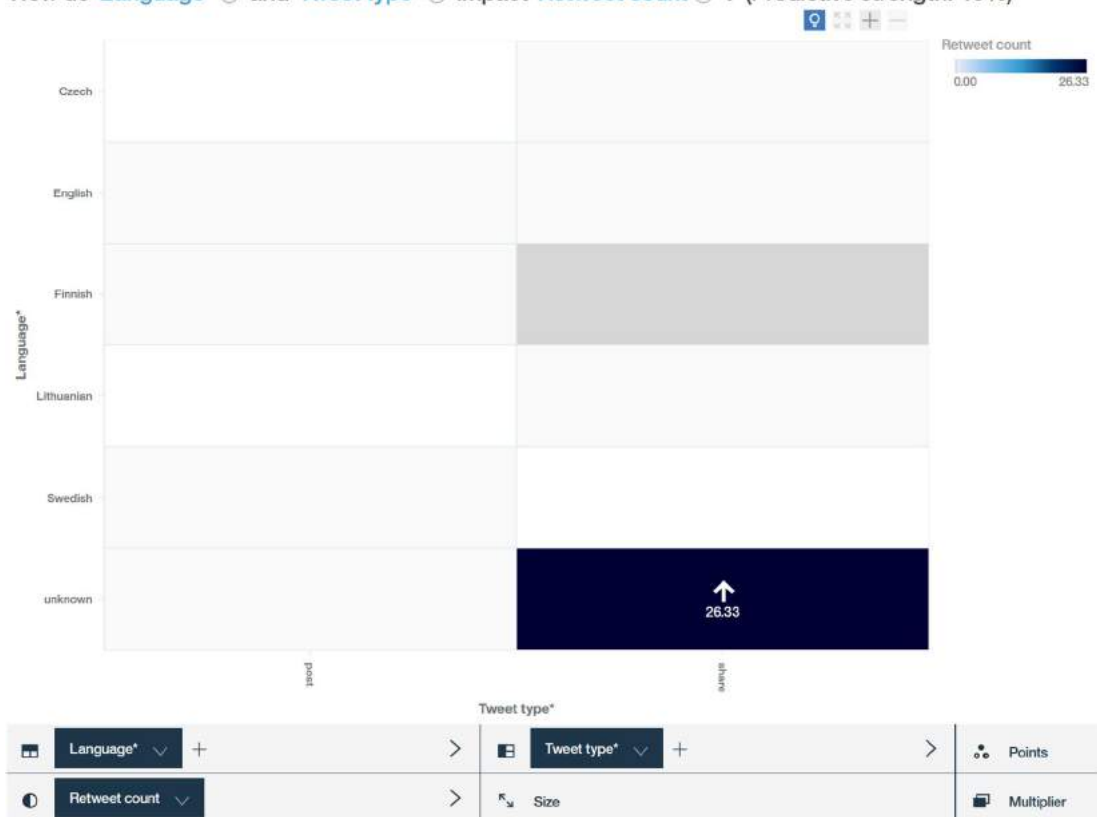


What drives Retweet count ?



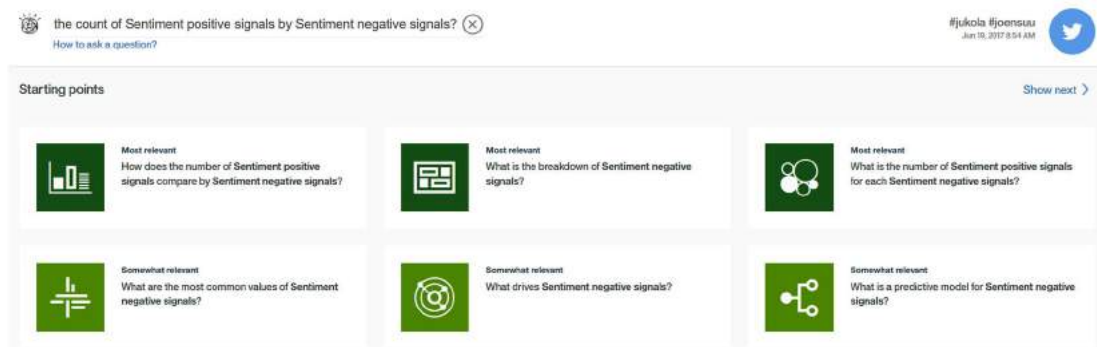
Kuva 100. Twiittien jakokertoihin vaikuttavat tekijät

How do Language\* and Tweet type\* impact Retweet count ? (Predictive strength: 46%)

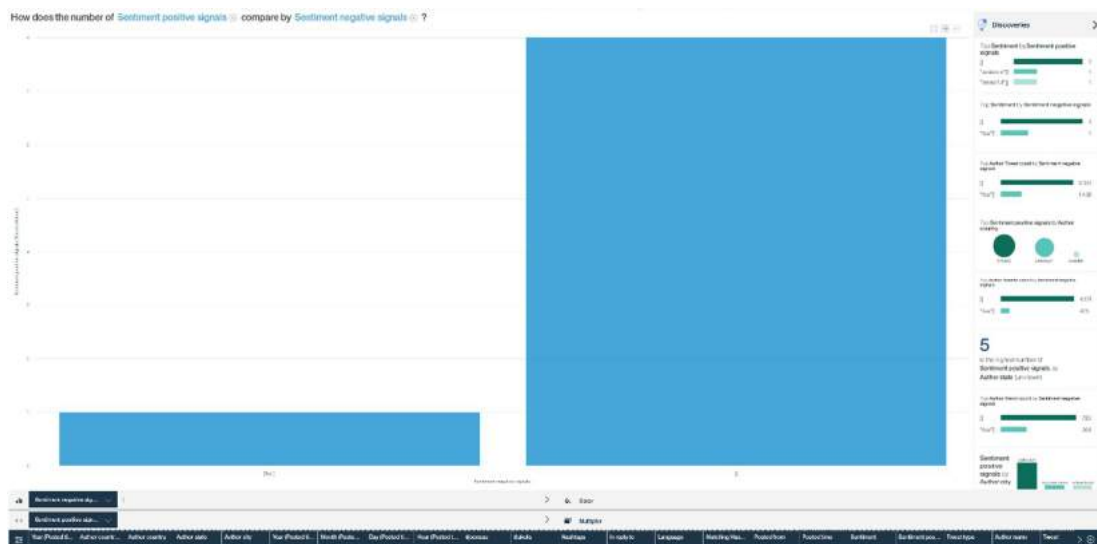


Kuva 101. Twiitin tyypin ja kielen vaikutus jakokertoihin

Vastataan kysymykseen "Paljonko oli asenteeltaa positiivisia twiittejä ja paljonko negatiivisia?" kysymyksellä "the count of Sentiment positive signals by Sentiment negative signals?" ja valitaan ensimmäinen näkemys What is the count of Sentiment positive signals by **Sentiment genative signals?** (Kuva 102). Kuvaajan (Kuva 103) mukaan positiivisen asenteen viestejä on kahdeksan kappaletta ja negatiivisia vain yksi. Löydetty yksittäinen negatiivinen sana on maksu (fee), joka näkyy vasemmanpuoleisen pylvään alla. Halutessaan positiivisia asenteita voi visualisoida tarkemmin esimerkiksi sanapilvellä tai uudella pylväsdiagrammilla, joka näyttää yleisimmät asenteet. Asenteiden määrä on tässä aineistossa pieni, sillä suurin osa viesteistä on suomen kielisiä.



Kuva 102. Kysymyksen pohjalta ehdotetut näkemykset

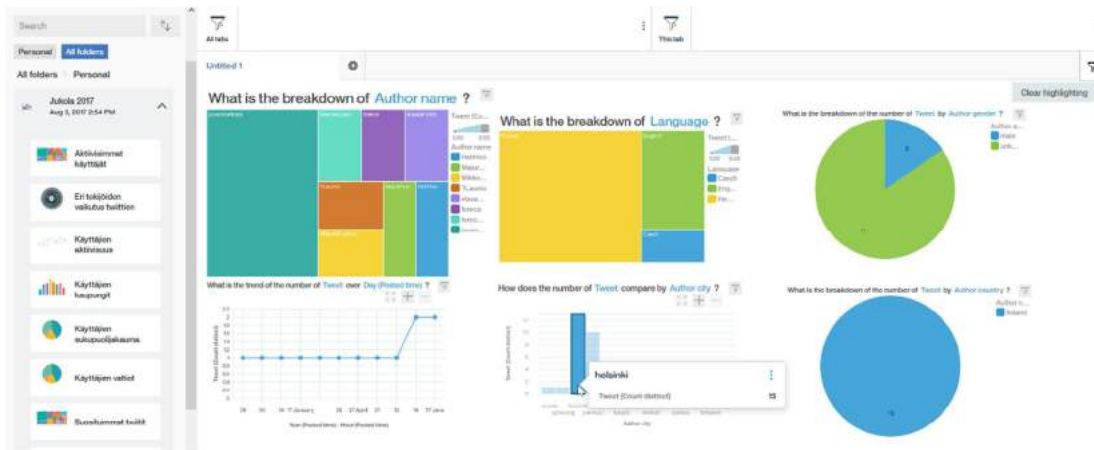


Kuva 103. Negatiivisten ja positiivisten asenteiden suhde

Tallennetaan luodut näkemykset painamalla yläreunan Save kuvaketta. Palataan aloitusnäkömään ja luodaan uusi tilannekuva Display välilehdestä painamalla New display (Kuva 104) ja lisätään luodut näkemykset (Kuva 105). Tallennetaan tilannekuva painamalla yläreunan Save kuvaketta.



Kuva 104. New display painike



Kuva 105. Tilannekuva Twitter datasta

### 3.4 FACEBOOK DATAN HANKKIMINEN

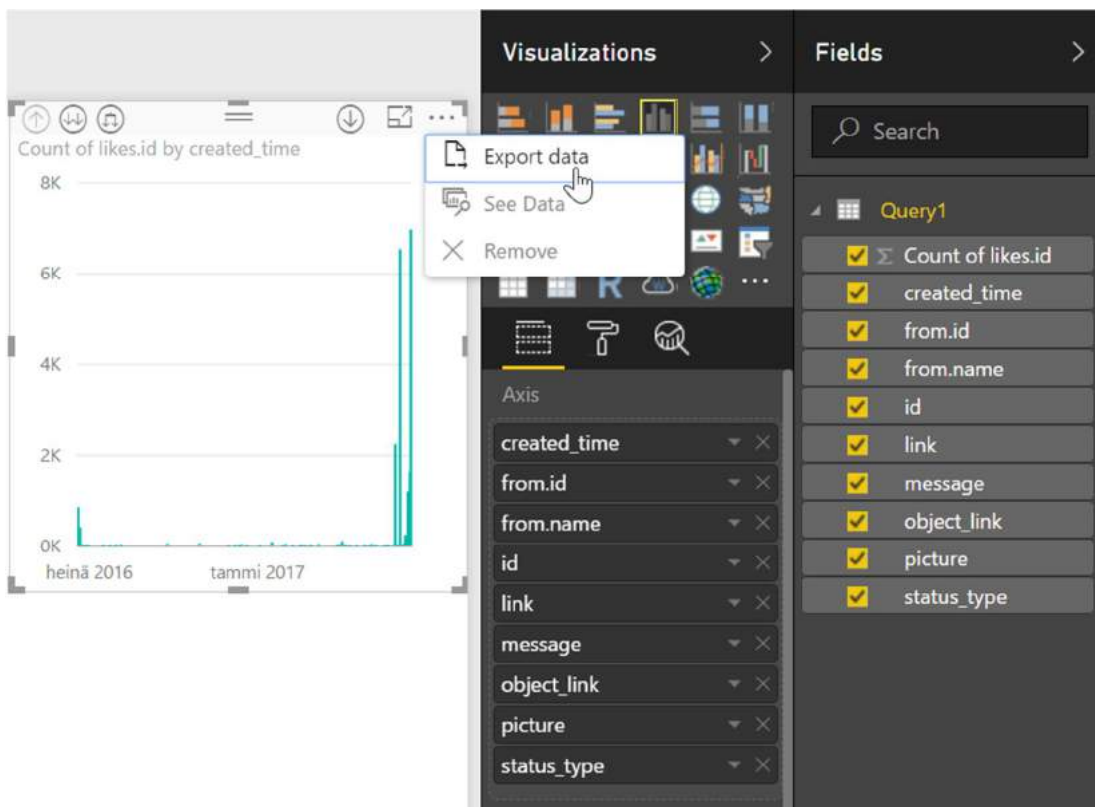
*Watson Analytics* ei tue Facebook datan hakemista suoraan *Facebookista*. Data voidaan kuitenkin hakea toisella työkalulla ja tuoda sen *Watson Analyticsiin*. Datajoukkona voidaan käyttää *Power BI*:lla haettua *Facebook* dataa. *Power BI* ei kuitenkaan tue suoraan datajoukon viemistä tiedostoon. Datajoukon viemiseksi taulukon tiedot tulee kopioida *Power BI*:n *Query Editorista* tai käyttää hyväksi visualisointien lataamista. *Power BI*:n työpöytäversioilla datajoukon saa ladattua helpoiten raporttityökalulla. Valitaan datajoukosta kaikki kentät, jonka jälkeen painetaan visualisoinnista ... > Export data (Kuva 106). Data tallennetaan .csv tiedostoon.

*Watson Analytics*in aloitusnäkymän *Data* osiossa painetaan *New data > Local file*, jonka jälkeen tuodaan tallennetun CSV-tiedoston työkaluun vetämällä sen ikkunaan tai selaamalla paikallisia tiedostoja *Browse* painikkeesta. *Watson Analytics* laskee datajoukon laaduksi 48 %. Avataan datajoukko tarkasteluun painamalla ... > Refine. Sarakkeet ovat viestin lähettäjä (from), luontiaika (created\_time), viestin tyyppi (status\_type), viestin id (id), objektin linkki (object\_link), tykkäysten määrä (likes), viestin sisältämä linkki (link) ja viestin sisältämä kuva (picture).

*Watson Analytics* ei pysty analysoimaan jäsentämätöntä dataa samalla tavalla kuin *Power BI*. Viestien tekstin analysoiminen ei ole mahdollista esimerkiksi sanapilvenä, koska viestisarake sisältää kokonaisia lauseita. *Watson Analytics* ei pysty myöskään enustamaan aikasarjadataa *Power BI*:n *Forecast* työkalun lailla.

Tarkastellaan datajoukon laatua valitsemalla oikealta *Data metrics*. Datajoukosta nähdään, että vienti *Power BI*:sta on onnistunut ja *Watson Analytics* on purkanut CSV-

tiedoston oikein sarakkeisiin (Kuva 107). Sarake `object_link` laskee huomattavasti datajoukon laatua, koska se sisälsi luettelon, joka tiedostoon tallennettaessa tallentui tekstinä. Painetaan oikeasta reunasta `Actions` ja poistetaan valinta sarakkeesta (Kuva 108). Tallennetaan muutokset datajoukkoon painamalla sivun yläreunasta `Save` painiketta. Piilotettu sarake nostaa datajoukon laatua 64 % asti (Kuva 109). Siirrytään näkemyksien luomiseen palaamalla aloitusnäkymän `Data` välilehteen ja painamalla datajoukkoa tai painamalla `Create discovery set` (Kuva 110) aloitusnäkymän `Discover` välilehdestä.



Kuva 106. Datajoukon vieminen csv tiedostoon Power BI:llä


Year (created_time)	Month (created_time)	Day (created_time)	created_time	from.id	id	link	message	object_link	picture	status_type	from.name	Count of likes.id
2017	June	15	2017-06-15 00:00:00	2500647935...	18268345792...	https://www.f...	Purnnit ei ole...	[Record]	https://sconte...	added_photos		6034
2017	June	6	2017-06-06 00:00:00	2500647935...	18268345792...	https://www.f...	Juoksuurto...	[Record]	https://sconte...	added_photos		6038
2017	June	1	2017-06-01 00:00:00	2500647935...	18268345792...	https://www.f...	Suunnistus on...	[Record]	https://sconte...	added_video		2256
2017	June	15	2017-06-15 00:00:00	3579366808...	18268345792...	https://www.f...	Kainuan Raati...	[Record]	https://sconte...	added_photos		1123
2017	June	17	2017-06-17 00:00:00	2521665246...	18268345792...	https://www.f...	Siene lähtövil...	[Record]	https://sconte...	added_video		583
2016	June	21	2016-06-21 00:00:00	1752395682...	18268345792...	https://www.f...	Photo by Julio...	[Record]	https://sconte...	added_photos		426


Kuva 107. Facebook datajoukon laatu

Columns
Search
<input checked="" type="checkbox"/> Year (created_time)
<input checked="" type="checkbox"/> Month (created_time)
<input checked="" type="checkbox"/> Day (created_time)
<input checked="" type="checkbox"/> created_time
<input checked="" type="checkbox"/> from.id
<input checked="" type="checkbox"/> id
<input checked="" type="checkbox"/> link
<input checked="" type="checkbox"/> message
<input type="checkbox"/> object_link

Kuva 108. Sarakkeen piilottaminen


+ New data

 Ask a question about y

 Personal


**Personal**

Last modified: Nov 7, 2016 1:20 PM




**Facebook Jukola**

Aug 3, 2017 6:26 PM



64% Quality



Kuva 109. Ladattu datajoukko



Kuva 110. New discovery set painike

## 3.5 NÄKEMYKSET FACEBOOK DATASTA

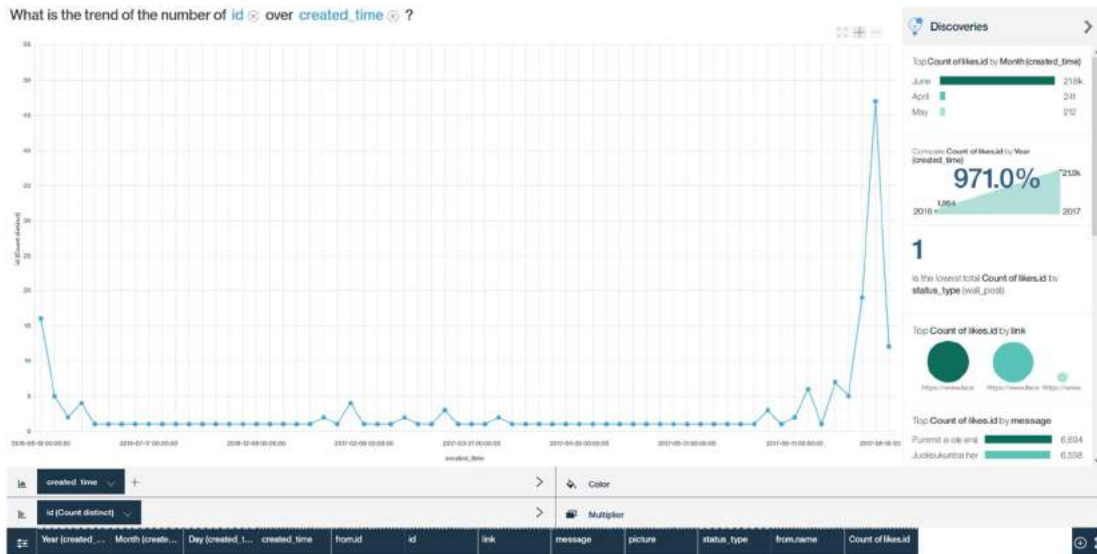
Tässä luvussa kuvataan näkemysten luominen *Watson Analytics*issä *Facebook* datalle. Luvussa 3.5.1 luodaan näkemyksiä manuaalisesti ja luvussa 3.5.2 pikanäkemysten ja kysymysten avulla.

### 3.5.1 Manuaalinen luominen

Visualisointityypit näkyvät *Create your own visualization* osiossa (Kuva 111). Vastataan kysymykseen "*Miten viestit jakautuivat määrällisesti tarkkailuajalle?*" vertaamalla viestien määrää päivämäärinä. Valitaan visualisoinnin tyyppi viivadiagrammi painamalla *Create your own visualization* osiosta *Line*. Vedetään hiiren osoittimella *created\_time* x-axis kenttään ja *id* y-axis kenttään. Kuvaajan (Kuva 112) mukaan käyttäjät lähettivät viestejä aktiivisimmin heti edellisen tapahtuman jälkeen sekä lähellä seuraavaa tapahtumaa. Heinäkuusta 2016 toukokuuhun 2017 sivulle on lisätty käyttäjien toimesta kuukausittain vain yksittäisiä viestejä. Sivulla vierailaan siis lähinnä vain tapahtumakuukautena. Avataan uusi välilehti seuraavalle näkemykselle yläreunan *Add a new tab* painikkeesta.

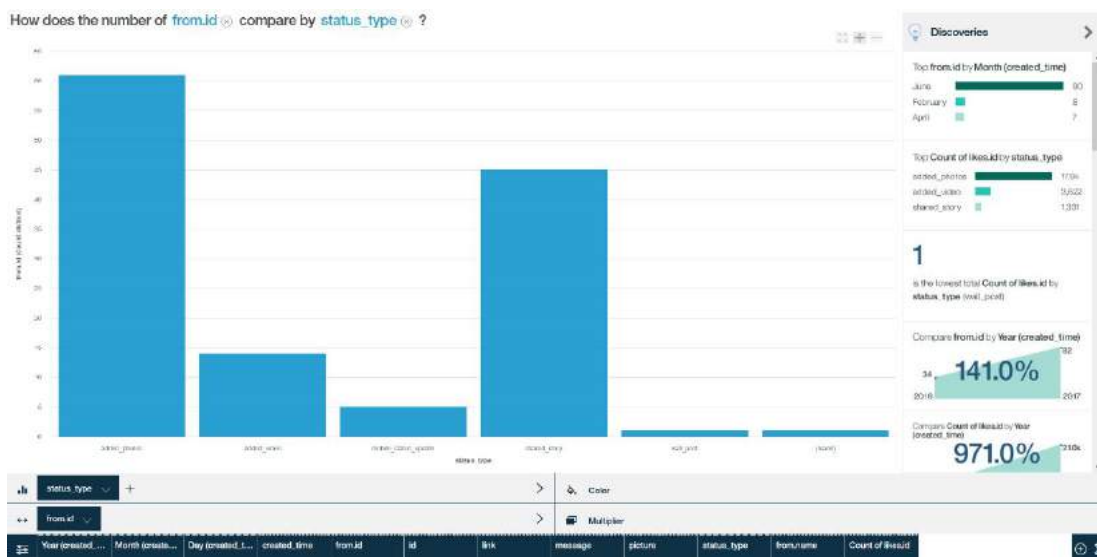


Kuva 111. Näkemysten manuaalinen luominen



Kuva 112. Uusien viestien määrä sivulla

Vastataan kysymykseen "Minkä tyyppisiä viestejä lisättiin?" vertaamalla kunkin käyttäjän lähettämiä viestityyppejä. Valitaan visualisoinnin tyyppiä pylväsdiagrammi painamalla `Create your own visualization` osiosta `Bar`. Vedetään hiiren osoittimella `status_type` `Bars` kenttään ja `from.id` `Length` kenttään. Kuvaajan (Kuva 113) mukaan käyttäjät ovat lisänneet eniten kuvaviestejä. 61 viestiä on tyybiltään kuvaviesti, 45 tekstiä ja 14 videoita. Lisäksi viisi viestiä on tyybiltään mobiilistatuspäivitys. Avataan uusi välilehti seuraavalle näkemykselle yläreunan `Add a new tab` painikkeesta.



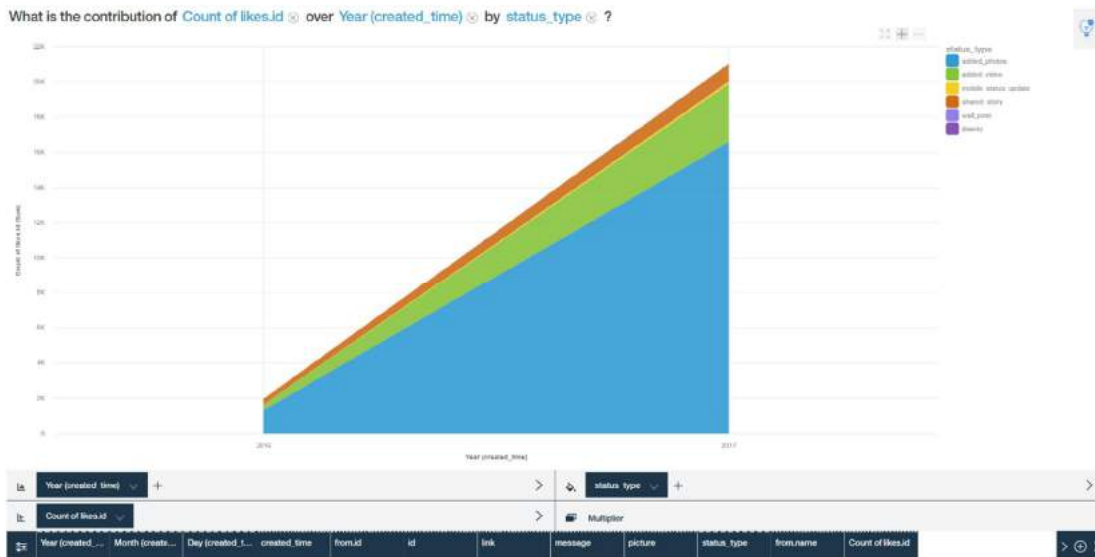
Kuva 113. Viestityyppien jakauma

### 3.5.2 Pikanäkemykset ja valmiit kysymykset

Pikanäkemykset näkyvät `Starting points` osiossa (Kuva 114). Valitaan kiinnostava pikanäkemyks `What is the contribution of Count of likes.id over Year (created_time) by status_type?`, joka on kuvakkeen mukaisesti tyypiltään aluekaavio. Kuvaajan (Kuva 115) mukaan kaikkia viestityyppejä on tykätty enemmän vuonna 2017 kuin vuonna 2016. Sosiaalisen median käyttäjät aktivoituvat vasta lähellä tapahtumaa. Avataan uusi välilehti seuraavalle näkemykselle yläreunan `Add a new tab` painikkeesta.



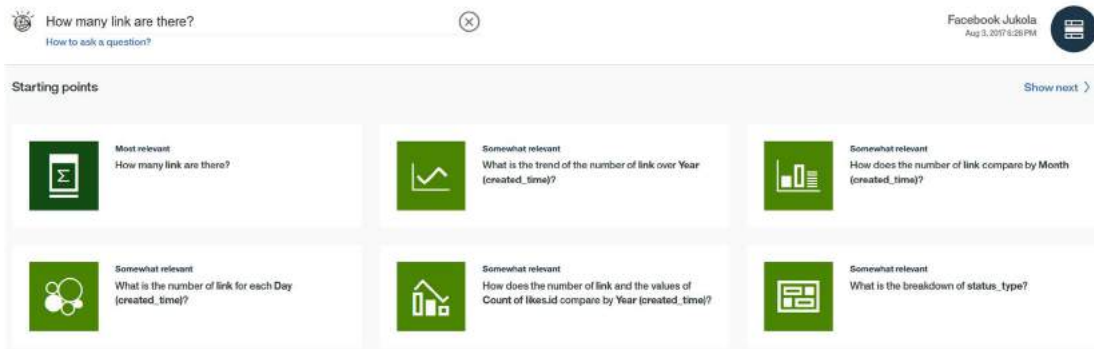
Kuva 114. Datajoukon pikanäkemykset



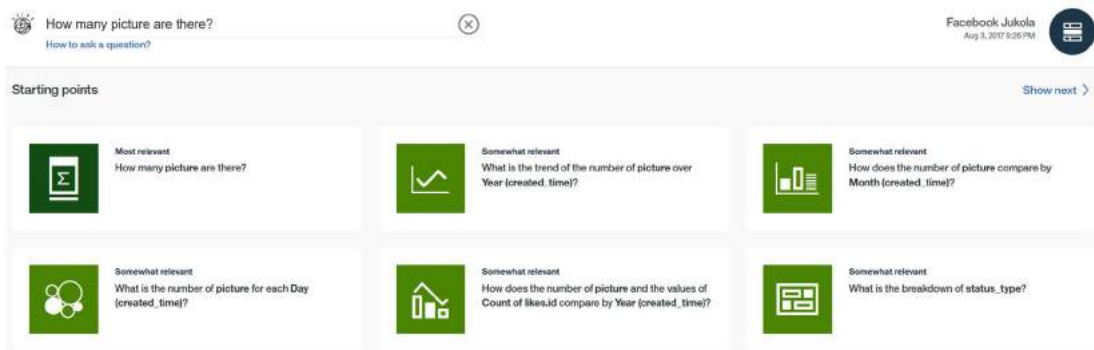
Kuva 115. Viestien tykkäuskerrat vuosina 2016 ja 2017

Vastataan kysymykseen "Kuinka moni viesti sisälsi linkkejä tai kuvia?" kysymyksillä "How many link are there?" (Kuva 116) ja "How many picture are there?" (Kuva 117) sekä valitsemalla näiden kysymysten tuottamat ensimmäiset näkemykset. Näkemyksen luku kertoo linkkien määrän (Kuva 118). Luku on kuitenkin yksilöllisten linkkien määrä. Vaihdetaan luku kokonaisnäkömään painamalla muuttujaa hiiren oikealla ja painamalla `Summarize > Count` (Kuva 119). Tehdään sama muutos myös kuvien määrälle. Kuvaajien (Kuva 120, Kuva 121) mukaan 188 viestistä 181 sisälsi ainakin yhden linkin ja 179 sisälsi ainakin yhden kuvan. Avataan uusi välilehti seuraavalle näkemykselle yläreunan `Add a new tab` painikkeesta.





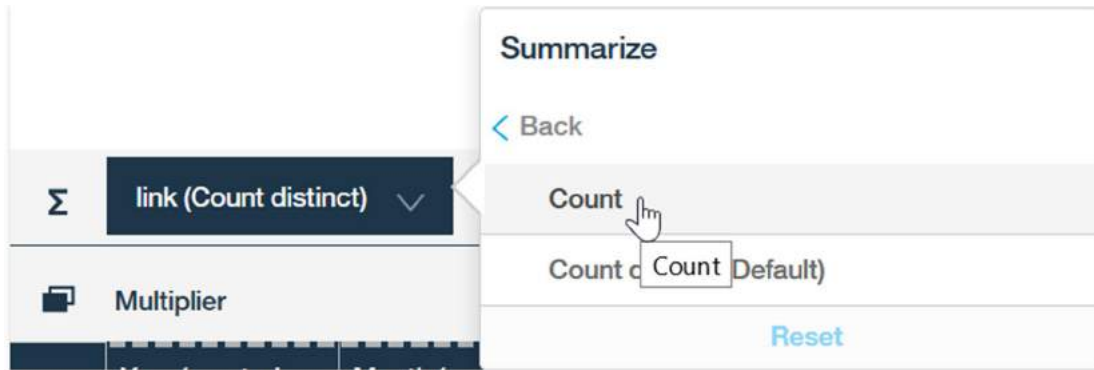
Kuva 116. Kysymyksen pohjalta ehdotetut näkemykset



Kuva 117. Kysymyksen pohjalta ehdotetut näkemykset



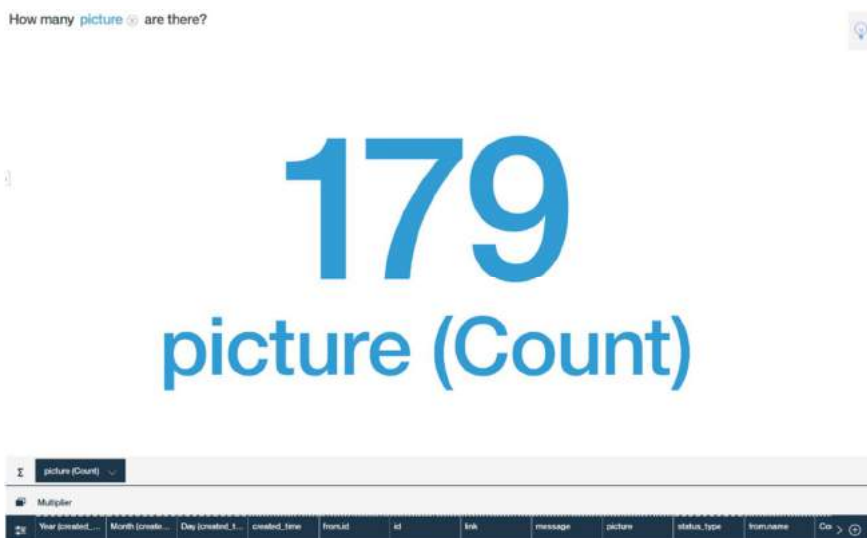
Kuva 118. Yksilöllisten linkkien määrä



Kuva 119. Yhteenvedon muuttaminen

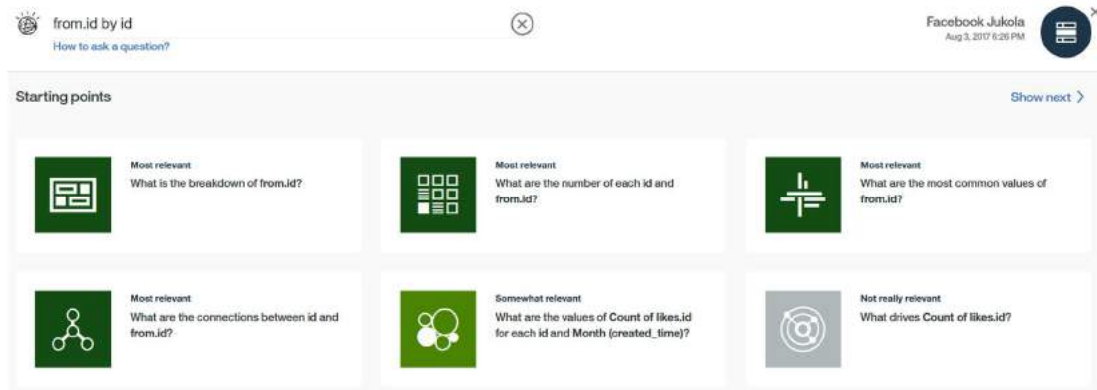


Kuva 120. Linkkien määrä

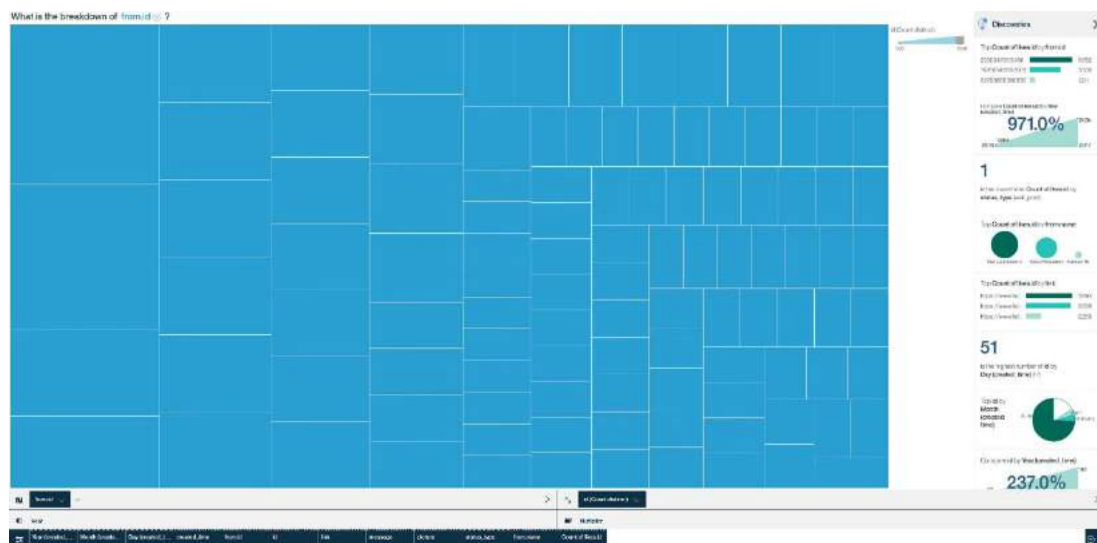


Kuva 121. Kuvien määrä

Vastataan kysymykseen "Montako viestiä yksittäiset henkilöt lisäsivät?" kysymyksellä "from.id by id" ja valitaan ensimmäinen näkemys what is the breakdown of from.id? (Kuva 122). Kuvaajan (Kuva 123) mukaan aktiivisin käyttäjä on lähettänyt 11 viestiä ja toiseksi aktiivisin 10. Suurin osa käyttäjistä on lähettänyt vain yhden viestin. Avataan uusi välilehti seuraavalle näkemykselle yläreunan Add a new tab painikkeesta.

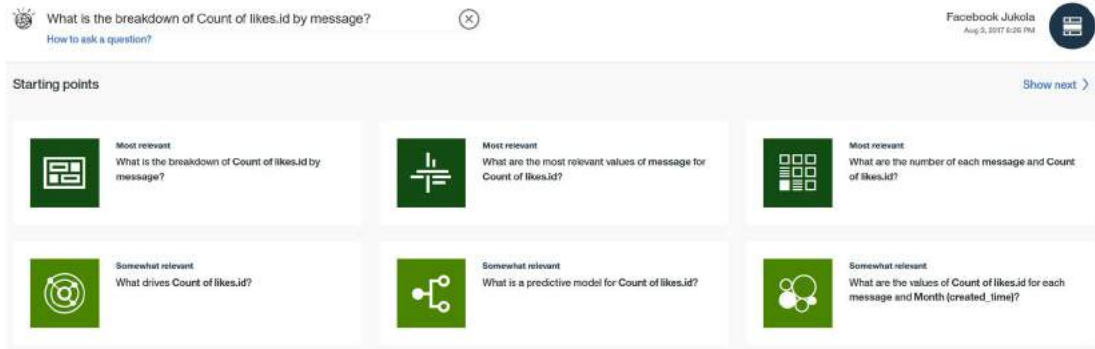


Kuva 122. Kysymyksen pohjalta ehdotetut näkemykset

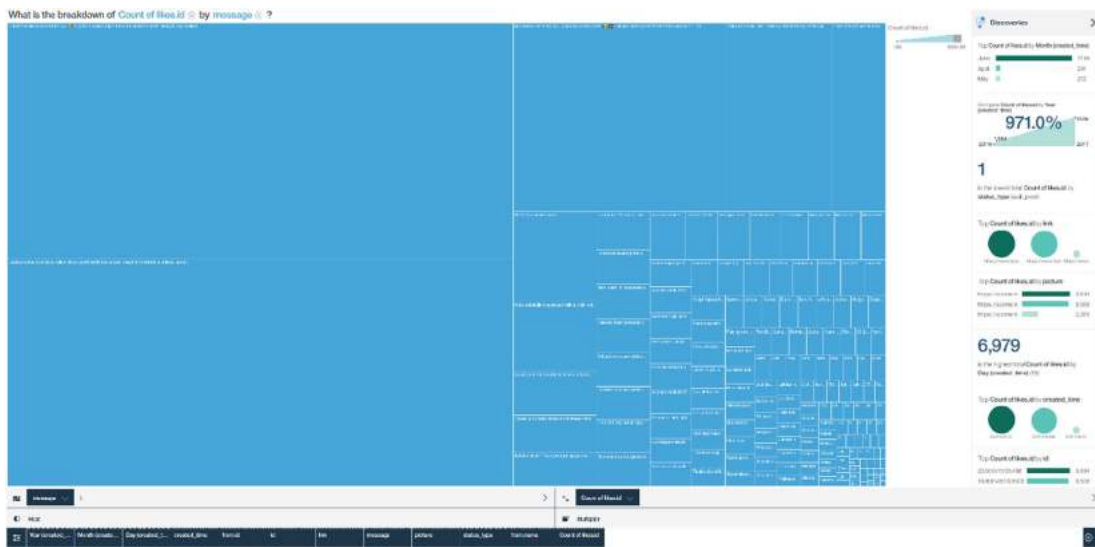


Kuva 123. Eri käyttäjien lähettämien viestien määrä (tunnisteet poistettu)

Vastataan kysymykseen "Mitkä olivat eniten tykättyt viestit?" kysymyksellä "What is the breakdown of Count of Likes.id by message?" ja valitaan sama ensimmäinen näkemys (Kuva 124). Kuvaajan (Kuva 125) mukaan kaksi viestiä on saanut huomattavasti enemmän tykkäyksiä kuin muut viestit.



Kuva 124. Kysymyksen pohjalta ehdotetut näkemykset

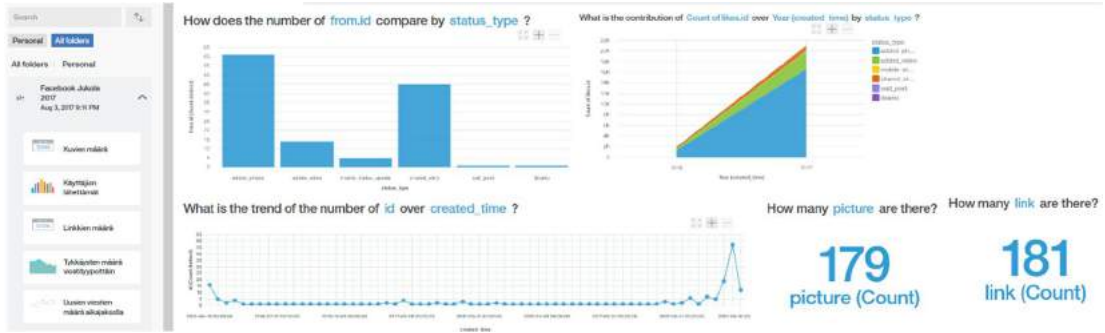


Kuva 125. Tykättyimmät viestit

Tallennetaan luodut näkemykset painamalla yläreunan Save kuvaketta. Palataan aloitusnäkömään ja luodaan uusi tilannekuva Display välilehdestä painamalla New display (Kuva 126) ja lisätään luodut näkemykset (Kuva 127). Tallennetaan tilannekuva painamalla yläreunan Save kuvaketta.



Kuva 126. New display painike



Kuva 127. Tilannekuva Facebook datasta

# 4 Googlen analytiikkatyökalut

Google tarjoaa eri tasoisia ilmaisia palveluita verkkoliikenteen seuraamiseen kuten *AdWords* (Luku 4.1), *Trends* (Luku 4.2) ja *Analytics* (Luku 4.3), joiden ilmaiseksi käytettävä analytiikka on esitelty tässä luvussa.

## 4.1 GOOGLE ADWORDS

*Google AdWords*<sup>30</sup> on palvelu, joka on suunniteltu verkkosivun mainoskampanjan toteuttamiseen *Googlen* hakukoneen kautta. Palvelua voi kuitenkin käyttää myös tietyssä määrin analytiikkaan, sillä se tarjoaa tietoja yleisistä hakutermeistä, joilla käyttäjät ovat löytäneet verkkosivulle.

*Google AdWordsin* avainsanoja selataksaan uuden tulee sivulla aloittaa uuden mainoskampanjan luominen (Kuva 128). Luomisprosessia ei kuitenkaan tarvitse viedä loppuun asti, joten se täysin ilmaista. Kampanjan luomiseen tarvitaan sähköpostiosoite ja seurattavan verkkosivun osoite. Osoitteella [jukola.com/2017](http://jukola.com/2017) voimme seurata liikennettä hakukoneesta vuoden 2017 tapahtuman sivuille. Painamalla **Jatka** päästään seuraavalle sivulle, jossa avainsanoja voi selata (Kuva 129).

Avainsanalistassa näkyy suosittuja sivustoon liittyviä avainsanoja, joita käyttäjät ovat hakukoneella etsineet. Suosituimpia hakuja olivat muun muassa jukolan viesti (9900 kappaletta) ja jukola (4400 kappaletta). Avainsanoja voi myös lisätä itse alareunan kentästä Anna lisää avainsanoja pilkulla eroteltuina. Avainsanojen suosiossa on otettava huomioon, että ne eivät kuvasta liikennettä, joka on ohjautunut sivustolle. Esimerkiksi joensuu majoitus haun suosio määräytyy myös muiden kuin itse tapahtuman kävijöiden hauissa. Haut tulee määritellä tarkkuudella, joka sulkee pois kohdejoukkoon kuulumattomat kävijät. Tapahtuman majoituksesta kiinnostuneiden käyttäjien määrä voidaan saada tarkemmin selville esimerkiksi avainsanoilla jukola 2017 majoitus tai majoitus jukola 2017 (Kuva 130). Avainsanojen suosio näyttää, että vuoden 2017 tapahtuman majoitusta on yritetty hakea hakukoneella 60 kertaa.

---

<sup>30</sup> <https://adwords.google.com>

# Tervetuloa AdWordsiin

Autamme sinua liittymisessä ja mainonnan aloittamisessa muutaman vaiheen avulla. Oletko kokenut AdWords-käyttäjä? [Ohita opastettu määrittäminen.](#)

Mikä on sähköpostiosoitteesi?

sposti@esimerkki.fi

▲ Kirjoita sähköpostiosoitteesi yllä olevaan kenttään.

Mikä verkkosivustosi on?

http://www.jukola.com/2017/

- Kyllä**, haluan mainosteni tehokkuuden parantamiseen tarkoitettuja henkilökohtaisia vinkkejä ja suosituksia.
- Ei**, en halua mainosteni tehokkuuden parantamiseen tarkoitettuja henkilökohtaisia vinkkejä ja suosituksia.

Jatka



Suorita muutama vaihe, niin voit aloittaa mainostamisen Googlessa.

Kuva 128. Google AdWords kampanjan luominen

## Ensimmäinen kampanjasi

**Kampanjassa** keskitytään tiettyyn aiheeseen tai tuoteryhmään. Luo kampanja asettamalla budjetti, valitsemalla yleisö ja kirjoittamalla mainos. Muista, että vaihtoehtojen valinnasta ei veloiteta. Voit myös muokata valintoja milloin tahansa.

1. Päätä, paljonko rahaa käytät

**Budjettisi** Päiväbudjetin määrittäminen

2. Valitse kohdeyleisö

**Sijainnit** Suomi

**Verkostot** Hakuverkosto, Display-verkosto

### Avainsanat

Lisää noin 15–20 avainsanaa. ?

Nämä ovat hakutermejä, jotka voivat käynnistää mainoksesi näkymään hakutulosten vieressä.

Avainsana	Hau suosio ?	
jukola 2016	590	×
jukolan viesti	9900	×
jukola	4400	×
jukolan viesti 2016	210	×
lahti energia	6600	×
majoitus lahti	880	×
viesti	3600	×
lahden energia	590	×
lappee jukola	170	×

Anna lisää avainsanoja pikulla erotettuina. Lisää

Tallenna Peruuta

Kuva 129. Avainsanojen selaaminen



Kuva 130. Hakulauseiden suosion selvittäminen

## 4.2 GOOGLE TRENDS

*Google Trends*<sup>31</sup> on Googlen hakukoneen hakuja visualisoiva ilmainen palvelu. Palvelun avulla pystytään seuraamaan eri hakutermin suosiota ajan sekä alueiden mukaan (Kuva 131). Syöttämällä yläreunan kenttään `Explore topics` hakutermin `jukola` saadaan tietoja Google hauista, jotka sisälsivät termin `jukola` (Kuva 132).

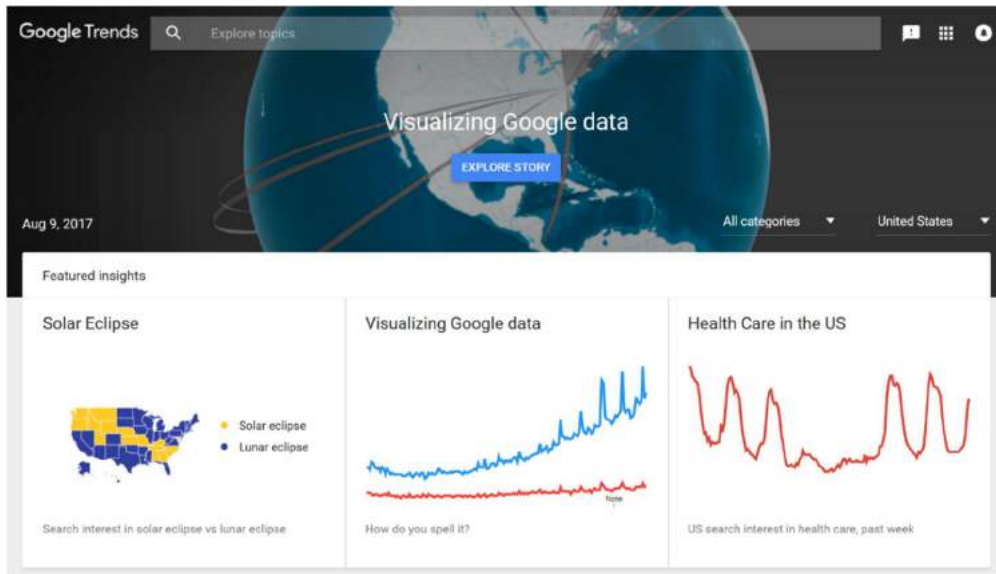
Hakutermin kuvaajien mukaan jukolaa on haettu eniten tapahtuman aikaan. Piikki hauissa on tapahtunut juuri tapahtuman alussa ja suurin hakumäärä on sijoittunut ajanjaksolle 11.-17. kesäkuuta. Hakumääriä mitataan prosentteina 100 % ollessa tarkkailujan suurin hakumäärä. Hakuja termillä on suoritettu kiinnostus alueittain kartan mukaisesti yksinomaan Suomesta. Aiheeseen liittyvää osio näyttää, että jukolaa hakeneita käyttäjiä kiinnostivat myös muun muassa Joensuu ja Suunnistus. Kohdassa Aiheeseen liittyvät kyselyt näytetään muut hakutermit, joilla käyttäjät hakivat tietoa tapahtumasta.

Haun tuloksia voi rajata yläreunan palkista valtioiden, ajan, luokkien ja haun kohteiden perusteella. Useampia hakutermejä voi myös verrata keskenään lisäämällä uusia termejä yläreunan kohdasta `vertaa`. Lisätään kaksi uutta termiä `joensuu` ja `joensuu majoitus` (Kuva 133). Vertailtavista hakumääristä nähdään, että joensuun kiinnostus ei noussut merkittävästi tapahtuman aikana ja joensuun majoitus hakujen määrä pystyi erittäin alhaisena myös tapahtuman aikana.

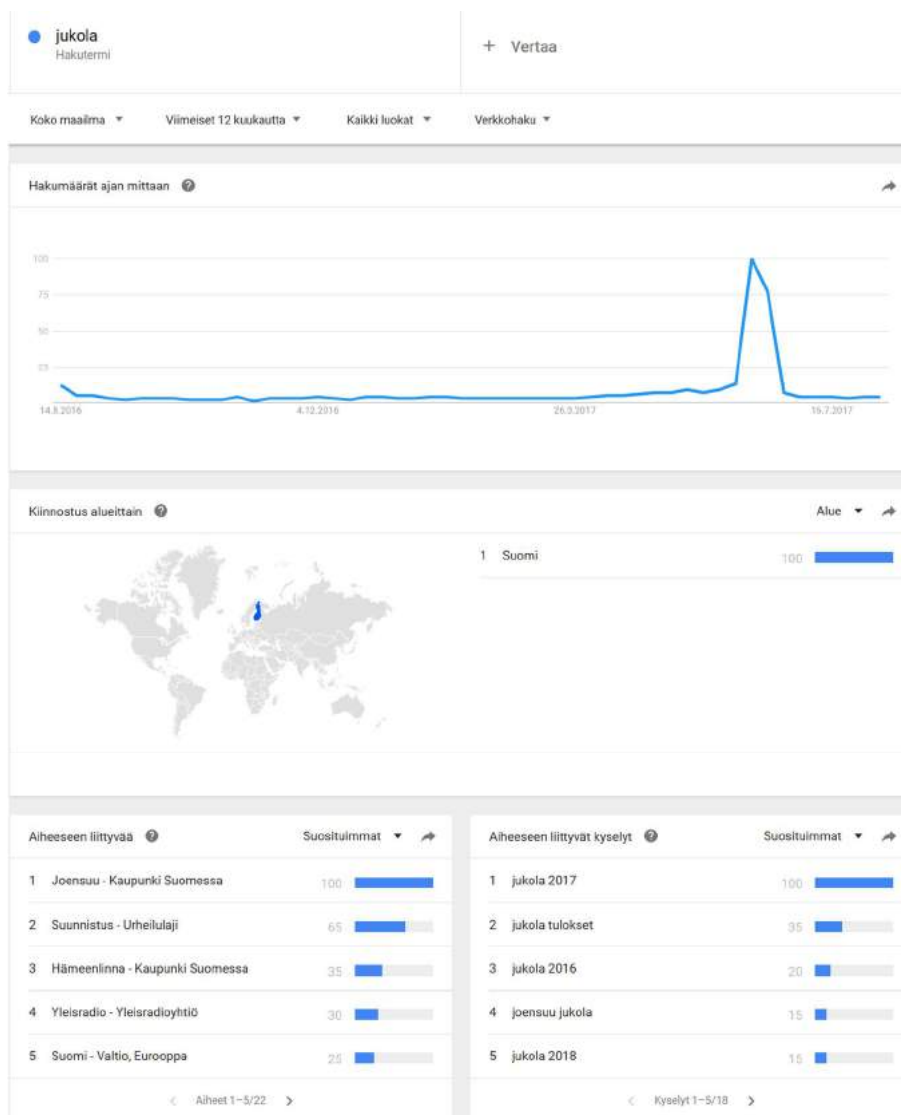
---

<sup>31</sup> <https://trends.google.com>

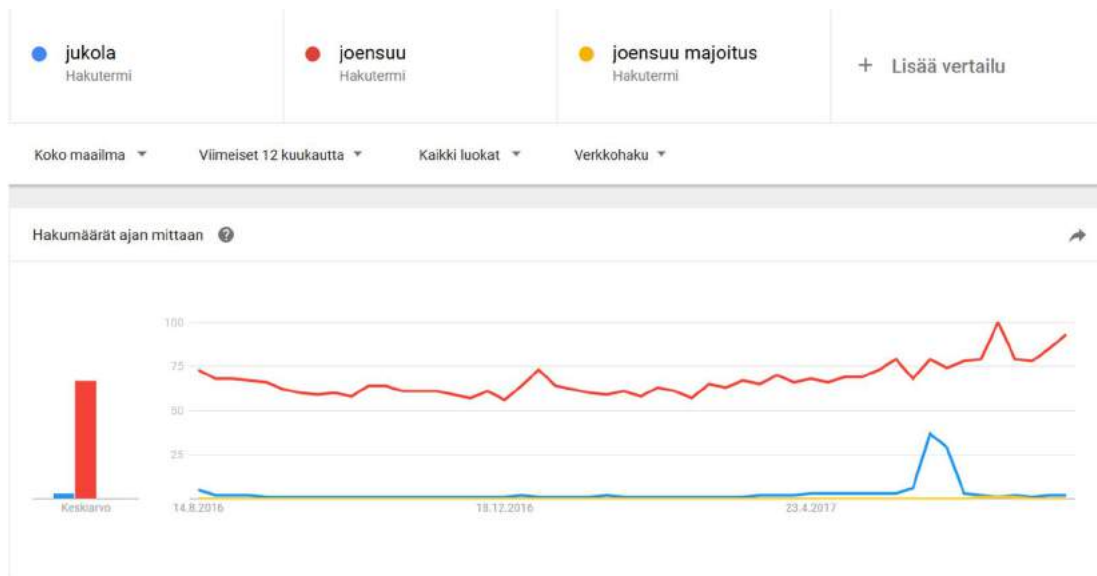




Kuva 131. Google Trends



Kuva 132. Hakutermin jukola tulokset



Kuva 133. Hakutermien vertailu

#### 4.3 GOOGLE ANALYTICS

Google tarjoaa myös *Freemium* mallista analytiikkapalvelua *Google Analytics*<sup>32</sup>. Palvelun käyttäminen on ilmaista ilman aikarajoitusta, mutta suuremman verkkoliikennemäärän analysoimisesta tulee maksaa. Palvelu toimii upottamalla omistamalleen verkkosivulle seurantakoodin, joka tarkkailee sivun kävijöitä (Kuva 134). *Google Analytics* pystyy raportoimaan tietoja esimerkiksi hakusanoista tai tavasta, joilla käyttäjät päätyivät sivulle sekä sosiaalisista verkostoista, jotka ohjasivat sivustolle. Näkemykset esitetään käyttäjälle tilannekuvan muodossa (Kuva 135).



Kuva 134. Google Analytics

<sup>32</sup><https://analytics.google.com>



Kuva 135. Esimerkki Google Analytics näkemyksistä

## 5 Futusomen palvelu

*Futusome*<sup>33</sup> on suomalainen yritys, joka on erikoistunut suomalaisen avoimen datan keräämiseen, jalostamiseen ja hyödyntämiseen. *Futusomen* kattavasta data-arkistosta voit selvittää mitä ja milloin suomalaiset kuluttajat ja yritykset viestivät sosiaalisessa mediassa.

*Futusomen SomeTutkija* työkalulla löytää somen sisällöt kattavasti mm. *Facebookista*, *Twitteristä*, *Instagramista*, *Pinterestistä*, keskustelupalstoilta, blogeista ja uutiskommenteista. *SomeTutkijassa* on kattava sosiaalisen median historiadata vuoden 2015 alusta tähän päivään ja hakuja voi tehdä *SomeTutkijassa* rajattomasti erittäin monipuolisilla hakutekijöillä.

KoDa- projektissa *Futusome* toimii sosiaalisen median datatoimittajana. *Futusomen* asiakkaat käyttävät sosiaalisen median sisältöä monipuolisesti apuna mm. asiakasymmärryksen parantamiseen, kilpailijoiden benchmarkkaukseen, kampanjan ajoituksen parantamiseen, muutoksen vertailututkimukseen ja aihepiiritutkimukseen. *Futusomen* asiakkaina on yrityksiä ja tutkimuslaitoksia, jotka välittävät omasta ja asiakkaidensa kilpailukyvystä ja haluavat käyttää parasta saatavilla olevaa dataa päätöksiensä tueksi.

Esimerkiksi haku "jukolanviesti" tuotti ajalta 1.1.2015 - 9.8.2017 yhteensä 9432 tulosta sosiaalisesta mediasta. Käyttöliittymästä näkee suoraan suosituimmat sivustot (Kuva 136), yleiset sanat (Kuva 137) ja adjektiivit (Kuva 138), hashtagit (Kuva 139) ja viestien ajallinen jakauma (Kuva 140) vuoden 2015 alusta alkaen. *SomeTutkijalla* haku tehdään luonnollisella kielellä eli yksinkertaisesti hakemalla sanan perusmuodolla. Datan laatu näkyy Suomen kielen käsittelyssä: sisällöstä tunnistetaan kieli, löydetään taivutusmuodot, yhdyssanat, sanaluokat, linkit, otsikot jne. *SomeTutkija* toimii selaimessa tai API-rajapinnan kautta tai datan voi tilata excel/json/csv -muodoissa (Kuva 141). *Futusomen* keräämä somen historiadata ulottuu jopa 10-vuoden taakse ja myös näistä voi tilata datadumppeja.

---

<sup>33</sup> <https://www.futusome.com/>

**9 432 tulosta  
sosiaalisesta mediasta**

**Lähdetyypit**

Twitter	9003
Instagram	294
Facebook	79
Keskustelupalsta	39
Blogit	14
Uutiskommentti	2
Google+	1

**10 SUOSITUINTA SIVUSTOA**

twitter.com	9000
www.instagram.com	205
instagram.com	89
www.facebook.com	79
keskustelu.suomi24.fi	29
www.demi.fi	5
www.fitoona.com	4
ask.fm	2
twitter.com/heituk	2
vaihtoaskelhyppy.blogspot.com	1

Kuva 136. jukolanviesti eri kanavissa

**VIESTIEN SISÄLTÖÄ KUVAAVIA YLEISIÄ SANOJA**

jukolanviesti	7895
jukola	4010
suunnistus	2178
suunnistua	1806
joensuu	837
viesti	682
venla	552
rasti	547
joukkue	519
jukolatiimi	511
tänään	482
the	474
orienteering	461
kisa	454
mukaan	436
viestiä	389
urheilu	300
suomi100	295
kesä	291
http	291
kiitos	273
viestin	273
osuus	269
kiito	265
tunnelma	260
lähtö	256
hieno	251
for	250
jukola'n	246
notokpikssk	244

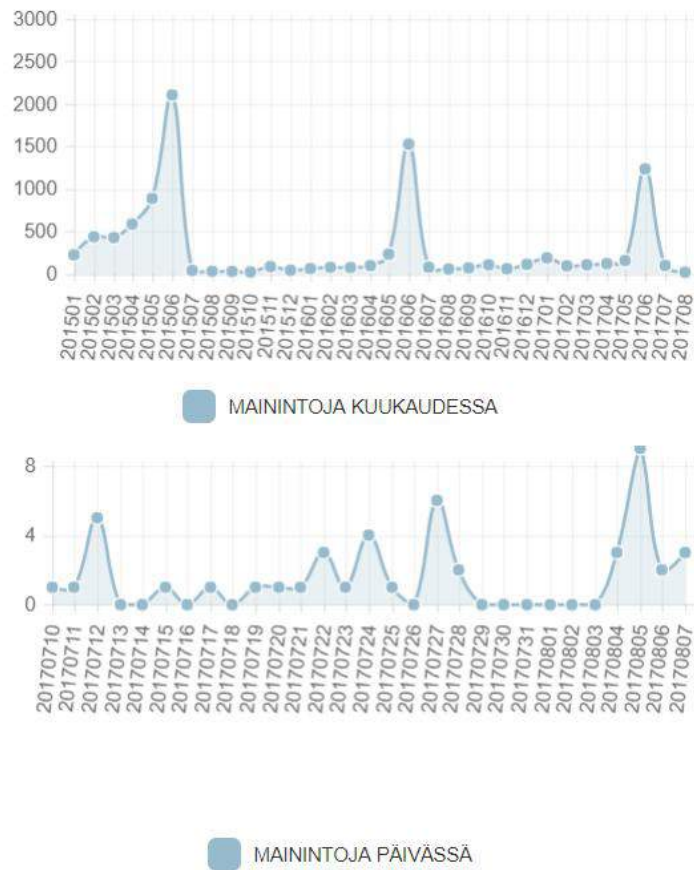
Kuva 137. jukolanviesti-viestien yleisimmät sanat

YLEISIMMÄT VIESTEISSÄ KÄYTETYT ADJEKTIIVIT	
paras	251
hieno	248
kova	130
valmis	102
upea	102
julkaistu	101
mahtava	98
viimeinen	97
suora	88
tuleva	88
kaunis	74
huikea	66
mukava	63
parkka	55
loistava	54
lämmin	52
ilmoittautunut	51
ajankohtainen	46
toivottava	44
rankka	42
matti	40
edullinen	39
virallinen	39
tuttu	38
avattu	38
aito	37
fyysinen	35
merkittävä	32
erinomainen	31
aitoa	30

Kuva 138. jukolanviesti-viestien yleisimmät adjektiivit

10 SUOSITUINTA #HASHTAGIA	
jukola	2552
suunnistus	1653
Jukola	1272
jukolanviesti	859
joensuu	681
orienteering	376
YhteinenJuttu	212
Suomi100	189
Suunnistus	160
JukolanViesti	155

Kuva 139. jukolaviestien 10 yleisintä hastagia



Kuva 140. jukolanviesti-viestien julkaisujakaumat

### Datan lataus

Tässä voit ladata koostetietoa tai viestidataa juuri tekemästasi hausta. Kun olet vahvistanut tilauksen, saat sähköpostiisi linkin datan latausta varten.

Kurkkaa tästä esimerkkejä sisällöstä ja eri tiedostomuodoista:

Koostetiedot: [Excel](#)

Viestidatan toimitetaan eri tiedostomuodoissa: [Excel & JSON & CSV](#)

Viestit (423 €)  
 Koostetiedot (212 €)  
 Viestit ja koostetiedot (529 €)

Maksullinen lataus laskutetaan työkalun laskutuksen yhteydessä. Jos haluat tilauksestasi erillisen laskun, syötä tähän tieto poikkeavasta laskutustavasta.

[Vahvista tilaus](#)

Kuva 141. Futosome-datan lataus

Jukolan viestin viestien tarkkailuajanjakso on 19.6.2016-18.6.2017. Futusome-palvelun avulla saadaan hyvä yleiskuva eli lähdetyyppi ja suosituimmat sivustot (Kuva 142), viestien aikasarjajakauma (Kuva 143), yleiset sanat (Kuva 144) ja adjektiivit (Kuva 145) sekä hashtagit (Kuva 146) vuoden 2015 alusta alkaen



Kuva 142. jukolanviesti-viestit lähdetypeittäin tarkkailuajanjaksona





Kuva 143. jukolanviestin-viestien jakauma ajankohtasidonnainen jakauma tarkkailuajanjaksolla

VIESTIEN SISÄLTÖÄ KUVAAVIA YLEISIÄ SANOJA	
jukolanviesti	2116
jukola	1100
joensuu	741
suunnistus	524
suunnistua	415
viesti	261
suomi100	248
venla	220
jukola2017	190
eno	165
rasti	142
mukaan	130
viestiä	129
juhla	128
orienteering	124
joukkue	123
viestin	115
status	113
kisa	108
the	107
yle	99
kesä	96
juhla viesti	92
and	91
kiitos	89
kiilto	85
klo	81
urheilu	79
for	79
ilmoittautua	77

Kuva 144. jukolanviestin-viestien yleiset sanat tarkkailuajanjaksolla

YLEISIMMÄT VIESTEISSÄ KÄYTETYT ADJEKTIIVIT	
suuri	72
paras	69
hieno	66
suora	52
upea	45
kaunis	42
aito	34
julkaistu	33
rankka	31
mukava	30
merkittävä	28
fyysinen	27
vaihteleva	26
valmis	26
taidollinen	26
aitoa	26
suomalainen	24
viimeinen	23
nuori	23
sähköinen	22
mahtava	20
avattu	20
huikea	20
lämmin	19
parka	18
ainutlaatuinen	18
tärkeä	18
erinomainen	17
vaativa	15
haastava	14

Kuva 145. jukolanviesti-viestien adjektiivit tarkkailuajanjaksolla

10 SUOSITUINTA #HASHTAGIA	
jukola	877
joensuu	631
suunnistus	402
jukolanviesti	377
Jukola	192
Suomi100	142
Jukola2017	116
orientteering	96
jukola2017	79
suomi100	70

Kuva 146. jukolanviesti-viestien hastagit tarkkailuajanjaksolla

<b>10 SUOSITUINTA KIRJOITTAJAA</b>	
<b>JukolanViesti</b>	453
<b>juusometsala</b>	52
<b>KalevanRasti</b>	30
<b>HannuVarkoi</b>	28
<b>SuunnistusSSL</b>	25
<b>SarkelaUlla</b>	21
<b>mariapauliinaa</b>	20
<b>TeamSOSE</b>	18
<b>Jukolanviesti</b>	18
<b>HarriHollo</b>	17

Kuva 147. jukolanviesti-viestien kirjoittajat tarkkailuajanjaksolla

# 6 Viestien kuva- ja tekstisisällön analysointi

Tässä luvussa tarkastellaan kognitiivisten palvelujen (cognitive services) tarjoamia mahdollisuuksia. Tällä hetkellä kognitiivisia palveluja ja niitä hyödyntäviä sovelluksia on kymmenittäin. IBM<sup>34</sup> ja Microsoft<sup>35</sup> mahdollistavat kognitiivisten palvelujen kokeilun. Tämän luvun esitykset pohjautuvat Hotti (2016), Hotti ja Gain (2016), Hotti ja Lauronen (2017), Gain ja Hotti (2017) sekä Gain et al. (2017) julkaisuihin. Lähempään tarkasteluun on otettu neljä kognitiivista palvelua – kaksi kuvien käsittelyyn (Luku 6.1) ja kaksi tekstin arviointiin (Luku 6.2 ja Luku 6.3).

Taulukko 5. Kognitiiviset palvelut, syötteet ja tuotosesimerkit, taulukossa MS=Microsoft

Kognitiivinen palvelu	Syöte	Tuotosesimerkkejä
IBM Visual Recognition <sup>36</sup>	Kuva	Kuva 149
MS Computer Vision API <sup>37</sup>	Kuva	Kuva 150
IBM Tone Analyzer <sup>38</sup>	Teksti	Kuva 152, Kuva 153, Kuva 154, Kuva 155
IBM Personality Insights <sup>39</sup>	Teksti	Kuva 157, Kuva 158

## 6.1 KUVIEN SISÄLLÖN ANALYSOINTI

Kun tarkastellaan viesteihin liittyviä kuvia (esimerkiksi Kuva 148). Niin niihin liittyvät analysointitulokset voidaan yhdistää muuhun viestidataan. Tämän luvun tarkoitus on havainnollistaa, millaisia asioita kuvista voidaan analysoida.

---

<sup>34</sup> <https://www.ibm.com/watson/products-services/>

<sup>35</sup> <https://azure.microsoft.com/en-us/services/cognitive-services/?v=17.29>

<sup>36</sup> <https://www.ibm.com/watson/developercloud/doc/visual-recognition/index.html>

<sup>37</sup> <https://www.microsoft.com/cognitive-services/en-us/computer-vision-api/documentation>

<sup>38</sup> <https://www.ibm.com/watson/developercloud/doc/tone-analyzer/index.html>

<sup>39</sup> <https://www.ibm.com/watson/developercloud/doc/personality-insights/index.html>

IBM *Visual Recognition*<sup>40</sup>(Kuva 149) tunnistaa kuvan kasvoista iän ja sukupuolen (Faces) sekä havaitsee kuvassa näkyviä asioita (Classes) ja laskee tunnistuksille pisteytyksen (Score). Microsoft *Computer Vision API*<sup>41</sup> (Kuva 150) yksilöi kuvasta tunnisteita (Tags), jotka perustuvat yli 2000 tunnistettavissa olevaan objektiin, elävään olentoon, maisemaan ja toimenpiteisiin (recognizable objects, living beings, scenery, and actions). Microsoft *Computer Vision API* arvioi tunnisteiden luotettavuuden ("confidence"). Ominaisuuksiltaan Microsoft *Computer Vision API* on monipuolisempi kuin IBM *Visual Recognition*, sillä se esimerkiksi tunnistaa, onko kuvassa aikuis sisältöä (Adult Content) ja onko kuva turhan rohkea (Racy) sekä pisteyttää arviot. Jos kuvassa on kasvoja, niin silloin Microsoft *Computer Vision API* tunnistaa iän ja sukupuolen kertoo kasvojen (Faces) sijainnin. Microsoft *Computer Vision API* tuottaa myös kuvasta sanallisen kuvauksen ja arvioi kuvauksen luotettavuutta (Description => "text").



Kuva 148. Jaetun viestin <https://t.co/hv5GFhPCpA> kuva<sup>42</sup>

---

<sup>40</sup> <https://visual-recognition-demo.mybluemix.net/>

<sup>41</sup> <https://www.microsoft.com/cognitive-services/en-us/computer-vision-api>

<sup>42</sup> <http://www.jukola.com/2017/2016/10/28/maavoimien-ja-joensuu-jukolan-yhteistyosopimus-allekirjoitettu/>

Classes	Score	Faces	Score	Type Hierarchy
military chaplain	0.60	age 45 - 54	0.40	/person/military chaplain
person	0.89	male	0.95	/person/serviceman/staff sergeant
staff sergeant	0.58	age 35 - 44	0.40	/person/serviceman/commissioned military officer/lieutenant colonel
serviceman	0.81	male	0.98	/person/serviceman/military recruit
lieutenant colonel	0.51	age 45 - 54	0.38	/person/army engineer
commissioned military officer	0.55	male	0.99	/person/serviceman/soldier/reservist
military recruit	0.51	age 55 - 64	0.61	
army engineer	0.50	male	0.73	
reservist	0.50			
soldier	0.50			
light brown color	0.73			
reddish orange color	0.58			

Kuva 149. IBM Visual Recognition tulokset

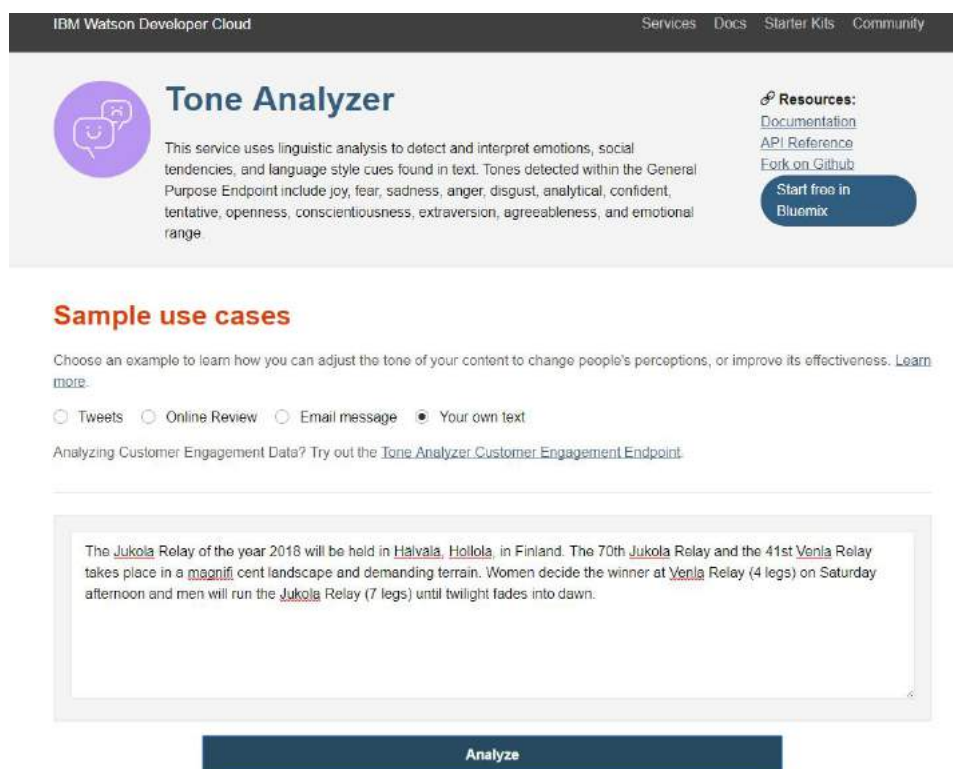
FEATURE NAME:	VALUE	FEATURE NAME:	VALUE	FEATURE NAME:	VALUE
Description	{ "tags": [ "person", "uniform", "indoor", "table", "man", "sitting", "clothing", "computer", "office", "laptop", "people", "group", "window", "food", "desk", "kitchen", "military", "talking", "cake", "counter", "eating", "woman", "room", "holding", "restaurant", "standing", "pizza", "shirt", "doing", "plate" ], "captions": [ { "text": "a group of people in uniform sitting at a table", "confidence": 0.9583432 } ] }	Image format	"Jpeg"	Dominant color background	■ "Grey"
Tags	[ { "name": "person", "confidence": 0.9997181 }, { "name": "military uniform", "confidence": 0.996079862 }, { "name": "indoor", "confidence": 0.994504631 }, { "name": "man", "confidence": 0.943489 }, { "name": "office", "confidence": 0.7679979 }, { "name": "people", "confidence": 0.6742893 }, { "name": "group", "confidence": 0.6243511 }, { "name": "working", "confidence": 0.542388856 } ]	Image dimensions	405 x 720	Dominant color foreground	■ "Black"
Adult content	false	Clip art type	0	Accent Color	■ #2D3A61
Adult score	0.0206490867	Line drawing type	0		
Racy	false	Black and white	false		
Racy score	0.0531840622				
Categories	[ { "name": "people_many", "score": 0.8515625 } ]				
Faces	[ { "age": 54, "gender": "Male", "faceRectangle": { "top": 88, "left": 543, "width": 118, "height": 118 } }, { "age": 46, "gender": "Male", "faceRectangle": { "top": 110, "left": 96, "width": 108, "height": 108 } }, { "age": 49, "gender": "Male", "faceRectangle": { "top": 82, "left": 232, "width": 43, "height": 43 } }, { "age": 53, "gender": "Female", "faceRectangle": { "top": 80, "left": 372, "width": 40, "height": 40 } } ]				

Kuva 150. Microsoft Computer Vision tulokset

## 6.2 LAUSEIDEN SÄVYT

*IBM Watson Analytics* analysoi *Twitter* viestin asenteet (sentiments) automaattisesti *Twitter* viestien latauksen yhteydessä. IBM:llä on myös kognitiivisia palveluja, joiden avulla voidaan analysoida esimerkiksi lauseiden tunteita, tyylejä ja persoonallisuustyppejä.

*IBM Tone Analyzer*<sup>43</sup> tunnistaa dokumentin sävyn eli tunteet (viha, inho, pelko, ilo ja surullisuus), kirjoitustyylin (analyttinen, luottavainen, epävarma) ja sosiaaliset taipumukset (avoimuus, tietoisuus, ulospäinsuuntautuneisuus, myöntövyys ja emotionaalinen vaihtelu) sekä arvioi sävyjä prosentuaalisesti. Esimerkiksi *Tone Analyzer* (Kuva 151) analysoi Jukolan esittelytekstin<sup>44</sup> kokonaisuudessaan (Kuva 152) iloiseksi (Kuva 153), luotettavaksi (Kuva 154) ja ekstroverttiseksi (Kuva 155), ja sävyjä voidaan lähemmin tarkastella lausetasolla.



IBM Watson Developer Cloud Services Docs Starter Kits Community

### Tone Analyzer

This service uses linguistic analysis to detect and interpret emotions, social tendencies, and language style cues found in text. Tones detected within the General Purpose Endpoint include joy, fear, sadness, anger, disgust, analytical, confident, tentative, openness, conscientiousness, extraversion, agreeableness, and emotional range.

**Resources:**  
[Documentation](#)  
[API Reference](#)  
[Fork on Github](#)  
[Start free in Bluemix](#)

#### Sample use cases

Choose an example to learn how you can adjust the tone of your content to change people's perceptions, or improve its effectiveness. [Learn more](#).

Tweets  Online Review  Email message  Your own text

Analyzing Customer Engagement Data? Try out the [Tone Analyzer Customer Engagement Endpoint](#).

The [Jukola](#) Relay of the year 2018 will be held in [Halvalla](#), [Hollola](#), in Finland. The 70th [Jukola](#) Relay and the 41st [Venla](#) Relay takes place in a [magnifi](#) cent landscape and demanding terrain. Women decide the winner at [Venla](#) Relay (4 legs) on Saturday afternoon and men will run the [Jukola](#) Relay (7 legs) until twilight fades into dawn.

Analyze

Kuva 151. Tone Analyzer<sup>45</sup>

<sup>43</sup> <https://tone-analyzer-demo.mybluemix.net/>

<sup>44</sup> <http://www.jukola.com/2018/files/2017/02/Jukolaesiteeng.pdf>

<sup>45</sup> <https://tone-analyzer-demo.mybluemix.net/>

## Document-level



Kuva 152. Jukolan viestin esittelytekstin yleisanalyysi

## Sentence-level

Identify sentences with stronger tones in context or sorted by score. Highlighted sentences indicate the likelihood of a tone present. If more than one tone is present, the stronger one is shown. Click on a sentence to see a breakdown of all tones.

**Tones**

**Emotion**

- Anger
- Disgust
- Fear
- Joy**
- Sadness

**Language**

- Analytical
- Confident
- Tentative

**Social Tendencies**

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Emotional Range

**In context**

Joy: Joy or happiness has shades of enjoyment, satisfaction and pleasure. There is a sense of well-being, inner peace, love, safety and contentment.

< .5   **.5 - .75**   > .75

None   Strong

The Jukola Relay of the year 2018 will be held in Hälväjä, Hollola, in Finland.

The 70th Jukola Relay and the 41st Venla Relay takes place in a magnificent landscape and demanding terrain.

**Women decide the winner at Venla Relay (4 legs) on Saturday afternoon and men will run the Jukola Relay (7 legs) until twilight fades into dawn.**

[Start Over](#)

Kuva 153. Iloisuus lausetasolla



## Sentence-level

Identify sentences with stronger tones in context or sorted by score. Highlighted sentences indicate the likelihood of a tone present. If more than one tone is present, the stronger one is shown. Click on a sentence to see a breakdown of all tones.

The screenshot shows a web interface for sentence-level analysis. On the left is a sidebar with categories: Tones, Emotion, Language, and Social Tendencies. Under 'Emotion', 'Confident' is selected. The main area, titled 'In context', shows a definition for 'Confident: A person's degree of certainty.' Below this is a filter bar with buttons for '< .5', '.5 - .75', and '> .75', with the middle one selected. A text box contains three sentences. The second sentence is highlighted in blue: 'Women decide the winner at Venla Relay (4 legs) on Saturday afternoon and men will run the Jukola Relay (7 legs) until twilight fades into dawn.'

Kuva 154. Luotettavuus lausetasolla

## Sentence-level

Identify sentences with stronger tones in context or sorted by score. Highlighted sentences indicate the likelihood of a tone present. If more than one tone is present, the stronger one is shown. Click on a sentence to see a breakdown of all tones.

The screenshot shows the same web interface as above, but with 'Extraversion' selected in the 'Emotion' sidebar. The definition in the main area is 'Extraversion: The tendency to seek stimulation in the company of others.' The filter bar is the same. The text box contains the same three sentences. The second sentence is highlighted in green: 'The 70th Jukola Relay and the 41st Venla Relay takes place in a magnificent landscape and demanding terrain.'

Kuva 155. Ekstroverttisyys lausetasolla

### 6.3 PERSOONALLISUUSPIIRTEET

IBM *Personality Insights*<sup>46</sup> tunnistaa tekstistä kirjoittajan 47 persoonallisuuden ominaispiirrettä - 30 piirrettä ovat Big Five -persoonallisuustyyppiin kuuluvia (Taulukko 6<sup>47</sup>), 12 piirrettä kuvastaa tarpeita (Taulukko 7<sup>48</sup>) ja viisi piirrettä arvoja (Taulukko 8<sup>49</sup>). Lisäksi palvelu tuottaa sanallisen yhteenvedon henkilön persoonallisuudesta ja havaitsee sanallisesti, mistä tekstin kirjoittaja mahdollisesti pitää ja mistä ei. Kognitiiviset sovellukset hyödyntävät *Personality Insights* palvelua muun muassa sopivan nettiterapeutin kohdentamisessa, sopivan oppilaitoksen suosittelemisessa, vaate ja muiden kulutushyödykkeiden suunnittelijoiden tuotteiden suosittelussa.

Tulevaisuudessa korostuu persoonallisuuspiirteiden (personality traits) tarkastelu ja jos käytettävissä on noin 100 sanaa. Esimerkiksi Jukolan historian tekstistä vuodelta 2009<sup>50</sup> saadaan tehtyä persoonallisuuspiirreanalyysi (Kuva 157). Yksittäisten piirteiden (Kuva 158) tuloksia voidaan tulkita.

---

<sup>46</sup> <https://personality-insights-livedemo.mybluemix.net/>

<sup>47</sup> <https://www.ibm.com/watson/developercloud/personality-insights.html>

<sup>48</sup> <http://www.ibm.com/watson/developercloud/doc/personality-insights/models.shtml#outputValue>

<sup>49</sup> <https://www.ibm.com/watson/developercloud/doc/personality-insights/models.shtml#outputNeeds>

<sup>50</sup> <http://www.jukola.com/en/jukolan-historia/>



# Personality Insights

Gain insight into how and why people think, act, and feel the way they do. This service applies linguistic analytics and personality theory to infer attributes from a person's unstructured text.

🔗 Resources:

[API Reference](#) [Documentation](#) [Fork on Github](#)

[Start free in Bluemix](#)

## Try the service

You need text written by the person whose personality you're interested in. It should contain words about every day experiences, thoughts, and responses.

You can play with the demo with as little as 100 words, but for a more accurate analysis, you need more words.

[Reset](#) | [Terms of use](#)

Tweets and Replies   **Body of Text**   Your Twitter Personality

---

Choose:   [2012 Debate - Barack Obama \(EN\)](#)   [Reflection - Gandhi \(EN\)](#)   [Michikusa - Natsume \(JA\)](#)

**Your own text**

The competition [centre](#) was located at [Mikkeli](#) hippodrome, close to the city [centre](#). The terrain had earlier served as training grounds for the Finnish Defense Forces. The relatively straightforward character thereof close to the hippodrome in particular caused some discussion prior to the competition, but concerns were near-universally dispelled afterwards. The special feature of [Jukola](#) Relay was the four-[kilometre](#)-long distance to the first control of the opening leg, which took the participants directly onto good terrain. 1106 [Venla](#)- and 1444 [Jukola](#) teams registered. Near 40.000 people followed [Jukola](#) start. [Ulricehamns](#) OK win [Venlas](#) and the [Venla](#) challenge cup forever. [Kristiansand](#) OK win [Jukola](#) after exciting competition.

Choose language:    English    Spanish    Arabic    Japanese   [Analyze](#)

Kuva 156. Personality Insights<sup>51</sup>

---

<sup>51</sup> <https://www.ibm.com/watson/services/personality-insights/>

## Personality Portrait

106 words analyzed: **Weak Analysis**

### Summary

You are skeptical, inner-directed and unconventional.

You are trusting of others: you believe the best in others and trust people easily. You are independent: you have a strong desire to have time to yourself. And you are authority-challenging: you prefer to challenge authority and traditional values to help bring about positive changes.

You are motivated to seek out experiences that provide a strong feeling of prestige.

You are relatively unconcerned with both helping others and taking pleasure in life. You think people can handle their own business without interference. And you prefer activities with a purpose greater than just personal enjoyment.

[How did we get this?](#)

You are likely to \_\_\_\_\_

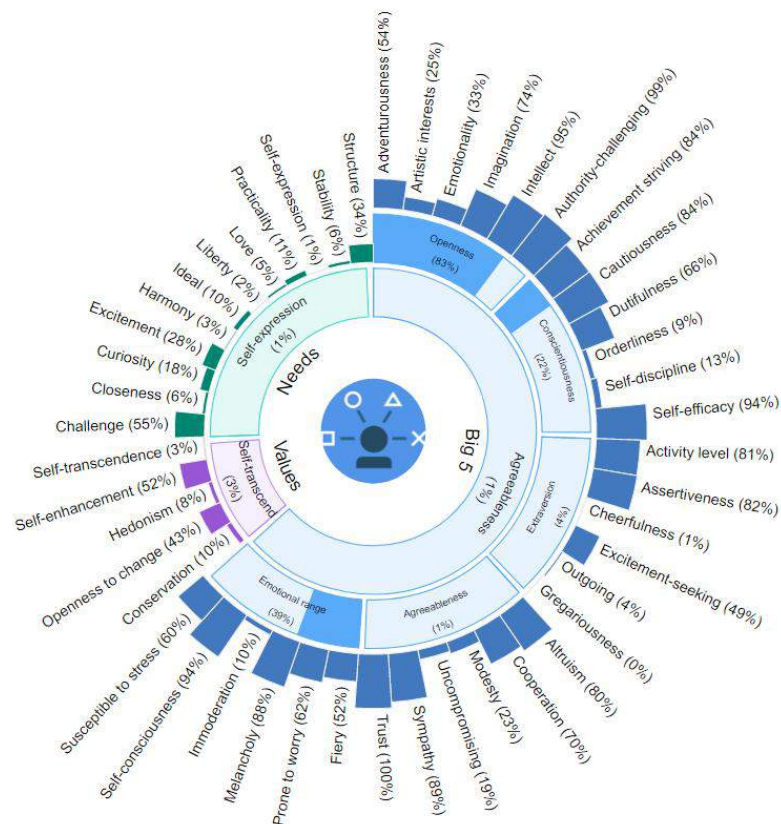
- like musical movies
- be sensitive to ownership cost when buying automobiles
- have experience playing music

You are unlikely to \_\_\_\_\_

- like country music
- be influenced by social media during product purchases
- be influenced by brand name when making product purchases

Kuva 157. Persoonallisuuspiirreanalyysin yhteenveto

## Sunburst Chart Visualization



Kuva 158. Aurinkopyrskäytys 47 persoonallisuuspiirteestä

Taulukko 6. 30 Big 5 piirrettä

Facet / trait	High explanation	Low explanation
Adventurousness	Adventurous: "You are eager to experience new things"	Consistent: "You enjoy familiar routines and prefer not to deviate from them"
Artistic interests	Appreciative of art: "You enjoy beauty and seek out creative experiences"	Unconcerned with art: "You are less concerned with artistic or creative activities than most people"
Emotionality	Emotionally aware: "You are aware of your feelings and how to express them"	Dispassionate: "You do not frequently think about or openly express your emotions"
Imagination	Imaginative: "You have a wild imagination"	Down-to-earth: "You prefer facts over fantasy"
Intellect	Philosophical: "You are open to and intrigued by new ideas and love to explore them"	Concrete: "You prefer dealing with the world as it is, rarely considering abstract ideas"
Authority-challenging	Authority-challenging: "You prefer to challenge authority and traditional values to effect change"	Respectful of authority: "You prefer following with tradition to maintain a sense of stability"
Achievement striving	Driven: "You set high goals for yourself and work hard to achieve them"	Content: "You are content with your level of accomplishment and do not feel the need to set ambitious goals"
Cautiousness	Deliberate: "You carefully think through decisions before making them"	Bold: "You would rather take action immediately than spend time deliberating making a decision"
Dutifulness	Dutiful: "You take rules and obligations seriously, even when they are inconvenient"	Carefree: "You do what you want, disregarding rules and obligations"
Orderliness	Organized: "You feel a strong need for structure in your life"	Unstructured: "You do not make a lot of time for organization in your daily life"
Self-discipline	Persistent: "You can tackle and stick with tough tasks"	Intermittent: "You have a hard time sticking with difficult tasks for a long period of time"
Self-efficacy	Self-assured: "You feel you have the ability to succeed in the tasks you set out to do"	Self-doubting: "You frequently doubt your ability to achieve your goals"
Activity level	Energetic: "You enjoy a fast-paced, busy schedule with many activities"	Laid-back: "You appreciate a relaxed pace in life"
Assertiveness	Assertive: "You tend to speak up and take charge of situations, and you are comfortable leading groups"	Demure: "You prefer to listen than to talk, especially in group situations"
Cheerfulness	Cheerful: "You are a joyful person and share that joy with the world"	Solemn: "You are generally serious and do not joke much"
Excitement-seeking	Excitement-seeking: "You are excited by taking risks and feel bored without lots of action going on"	Calm-seeking: "You prefer activities that are quiet, calm, and safe"
Outgoing	Outgoing: "You make friends easily and feel comfortable around other people"	Reserved: "You are a private person and do not let many people in"
Gregariousness	Sociable: "You enjoy being in the company of others"	Independent: "You have a strong desire to have time to yourself"
Altruism	Altruistic: "You feel fulfilled when helping others and will go out of your way to do so"	Self-focused: "You are more concerned with taking care of yourself than taking time for others"
Cooperation	Accommodating: "You are easy to please and try to avoid confrontation"	Contrary: "You do not shy away from contradicting others"
Modesty	Modest: "You are uncomfortable being the center of attention"	Proud: "You hold yourself in high regard and are satisfied with who you are"
Uncompromising	Uncompromising: "You think it is wrong to take advantage of others to get ahead"	Compromising: "You are comfortable using every trick in the book to get what you want"
Sympathy	Empathetic: "You feel what others feel and are compassionate toward them"	Hard-hearted: "You think people should generally rely more on themselves than on others"
Trust	Trusting of others: "You believe the best in others and trust people easily"	Cautious of others: "You are wary of other people's intentions and do not trust easily"
Fiery	Fiery: "You have a fiery temper, especially when things do not go your way"	Mild-tempered: "It takes a lot to get you angry"
Prone to worry	Prone to worry: "You tend to worry about things that might happen"	Self-assured: "You tend to feel calm and self-assured"
Melancholy	Melancholy: "You think quite often about the things you are unhappy about"	Content: "You are generally comfortable with yourself as you are"
Immoderation	Hedonistic: "You feel your desires strongly and are easily tempted by them"	Self-controlled: "You have control over your desires, which are not particularly intense"
Self-consciousness	Self-conscious: "You are sensitive about what others might be thinking of you"	Confident: "You are hard to embarrass and are self-confident most of the time"
Susceptible to stress	Susceptible to stress: "You are easily overwhelmed in stressful situations"	Calm under pressure: "You handle unexpected events calmly and effectively"

Taulukko 7. 5 Values piirrettä

Trait	High definitions
Conservation	"Emphasize self-restriction, order, and resistance to change"
Openness to change	"Emphasize independent action, thought, and feeling, as well as a readiness for new experiences"
Hedonism	"Seek pleasure and sensuous gratification"
Self-enhancement	"Seek personal success"
Self-transcendence	"Show concern for the welfare and interests of others"

Taulukko 8. 12 Needs piirrettä

Trait	High definitions
Challenge	"Have an urge to achieve, to succeed, and to take on challenges"
Closeness	"Relish being connected to family and setting up a home"
Curiosity	"Have a desire to discover, find out, and grow"
Excitement	"Want to get out there and live life, have upbeat emotions, and want to have fun"
Harmony	"Appreciate other people, their viewpoints, and their feelings"
Ideal	"Desire perfection and a sense of community"
Liberty	"Have a desire for fashion and new things, as well as the need for escape"
Love	"Enjoy social contact, whether one-to-one or one-to-many. Any brand that is involved in bringing people together taps this need"
Practicality	"Have a desire to get the job done, a desire for skill and efficiency, which can include physical expression and experience"
Self-expression	"Enjoy discovering and asserting . . . own identities"
Stability	"Seek equivalence in the physical world . . . favor the sensible, the tried and tested"
Structure	"Exhibit groundedness and a desire to hold things together . . . need things to be well organized and under control"

## Lähdeluettelo

Gain U., Hotti V. (2017): Tones and traits – experiments of text-based extractions with cognitive services. *Finnish Journal of eHealth and eWelfare*, 9(2-3), 82-94. DOI: <https://doi.org/10.23996/fjhw.61001>

Gain U., Hotti V., Lauronen H. (2017): Automation capabilities challenge work activities cognitively. *Tekoäly ja ihmistyön tulevaisuus, Futura 2/2017*, (ISSN 0785-5494).

Hotti V., Lauronen H. (2017): Kognitiivisten palvelujen ja kognitiivisten prosessien kohdentaminen. *Finnish Journal of eHealth and eWelfare*, 9(2-3), 121-135. DOI:<https://doi.org/10.23996/fjhw.61006>

Hotti V., Gain U. (2016): Exploitation and Exploration Underpin Business and Insights Underpin Business Analytics. In: Li H., Nykänen P., Suomi R., Wicramasinghe N., Widén G., Zhan M. (eds) Building Sustainable Health Ecosystems. WIS 2016. Communications in Computer and Information Science, Vol. 636. Springer, Cham, 223-237.

Hotti V. (2016): Tekoäly ja sisällön tulkinta haastavat Big Datan louhijat (Artificial intelligence and content interpretations challenge Big Data miners). *Big data – nykytila ja tulevaisuuden mahdollisuudet, Futura 2/2016*, 70-79.